

Australian Education Review

Using Data to Support Learning in Schools

Students, teachers, systems

Gabrielle Matters

First published 2006
by ACER Press
Australian Council *for* Educational Research Ltd
19 Prospect Hill Road, Camberwell, Victoria, 3124

Copyright © 2006 Australian Council for Educational Research

All rights reserved. Except under the conditions described in the *Copyright Act 1968* of Australia and subsequent amendments, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the written permission of the publishers.

Edited by Carolyn Glascodine
Typeset by Darren Jeffrey
Printed by Print Impressions

National Library of Australia Cataloguing-in-Publication data:

Matters, Gabrielle.
Using data to support learning in schools : students,
teachers, systems.

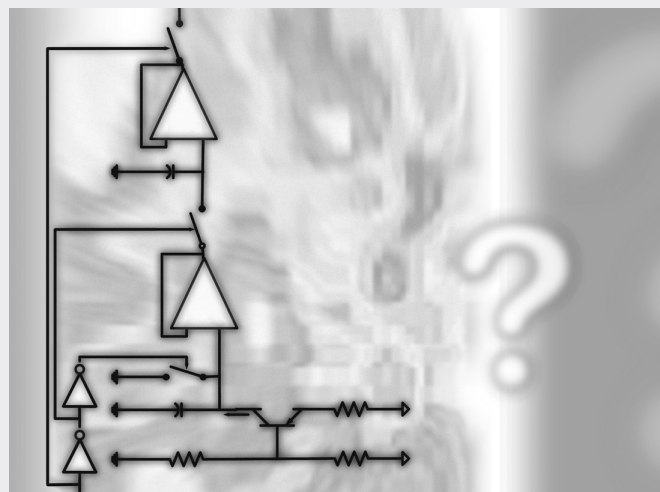
Bibliography.
ISBN 0 86431 692 5.

1. Teaching - Methodology. 2. Curriculum planning.
I. Title. (Series : Australian education review ; no. 49).

371.1

Visit our website: www.acer.edu.au

Foreword



In 'Using data to support learning', Gabrielle Matters envisions an educational system built around 'evidenced-based practice', the idea that decisions at all levels should be grounded in data. From this perspective, we are not only concerned with giving teachers the data they need to make more informed decisions about their students, but with all of the decision makers that constitute the educational system and all of the decisions they need to make to facilitate achievement.

Matters points out that the demand to improve decision making through data occurs because we are living in an age of accountability facilitated by new technology. But we should understand that accountability in education has not come about simply because we can collect and analyse data more easily. Accountability has become critical because today's globalised economy means that companies can invest anywhere that is politically stable and that has a skilled, productive workforce. Anywhere. To beat their competitors, companies must go to those locations that afford the best mix of skills and productivity. To keep jobs and to maintain current living standards, governments need to constantly improve the skill levels and productivity of their existing workforces. But to guarantee that future living standards are maintained, those governments must also ensure that today's students are educated to the highest achievement standards possible. And schools must be held accountable for that achievement if those standards are to be met.

For evidenced-based practice to have the maximum effect, we will need to make progress on several fronts. Matters notes that one essential front is in taking better advantage of research, and she identifies some of the questions and processes needing greater attention. A critical area largely ignored by test constructors has been research on the nature of proficiency in such domains as reading, writing, mathematics, and the sciences. Without building on that research, it is unlikely that our assessments will produce data that we can use effectively to improve student achievement or to enhance educational systems more generally.

Taking better advantage of this research means, first, defining each of the target domains not simply in terms of the contents that students need to master. It means defining and organising those domains in terms of:

- the strategies, processes and knowledge structures that are important to successful performance on the valued tasks that constitute the essence of the domain
- the features of valued tasks that call upon those strategies, processes and knowledge structures.

Conceptualising domain proficiency in this way is important because it should help us determine:

- what components of proficiency to assess
- what components to target for instruction
- what features of problems to manipulate to give diagnostic information or targeted instructional practice
- how to arrange instruction for maximum effect
- how to link proficiency components to government-stipulated competencies and qualifications.

Matters makes the case for focusing on student work. Thus, she leads us to a second front on which progress must be made: taking better advantage of new technology in our assessment devices and our learning tools. In virtually all service industries, workers use computers for at least some tasks. In knowledge-intensive jobs, workers use computers almost continuously. Going to university today means writing with a word processor, using the Internet to locate information, creating online presentations, and modelling problem situations with a spreadsheet. Not surprisingly, the computer is becoming a standard part of the academic toolkit for school-age students too. We will not be able to continue to make credible statements about how effectively students write, solve problems, or manage information if we base those statements solely on data gathered from paper assessments. We will need to find ways to measure performance effectively in the online environments in which students are doing their work – for example, their writing, problem solving, and information management.

Last, we'll need to make better use of advances in measurement science. Again, Matters' review prompts some thoughts about which of these advances might best support evidence-based practice. Of particular relevance are those advances that provide more powerful methods for making sense of student performance in complex assessment situations. Complex performance assessments like simulations are critical if we are to reflect the nature of domain proficiency faithfully. Educational measurement is, at its heart, about two things. The first thing is to characterise an individual's (or group's) domain proficiency in some way. That characterisation could be on an ordered scale or it could be into a diagnostic category. Second, measurement means providing an estimate of the uncertainty associated with that characterisation. Our current measurement models (like IRT and its derivatives) achieve those goals reasonably well when the assessment design presumes a large number of conditionally independent tasks, each tapping the same, single proficiency. But those same models are less effective when the assessment designs entail smaller numbers of extended problem-solving exercises which each measure multiple proficiencies.

So, for evidenced-based practice to have the maximum effect, we will need to make better use of advances in cognitive research, technology, and measurement science. However, a piecemeal approach to incorporating these advances won't suffice. The approach will need to be an integrated one. The work of Mislevy and his colleagues offers one promising path, especially for the case of complex performance assessments (Mislevy, Almond, & Lukas, 2003)¹. That approach, called 'Evidence-Centered Design', attempts to deal with some of the key issues raised by Matters, particularly the nature of evidence, data, and interpretations.

From Mislevy et al.'s perspective, effective assessment-based decision making depends on effective assessment design. Effective assessment design, in turn, depends upon a clear and complete explication of the claims one wishes to make on the basis of assessment, the evidence needed to support those claims, the student behaviours that will comprise that evidence, and the tasks needed to elicit those behaviours. That chain of reasoning, once established and documented, becomes part of the validity argument in support of the assessment. Further, reversing the chain allows inferences to be made from data. That is, observing a student behaviour provides partial evidence for a claim about that student's domain proficiency.

¹ Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to Evidence-Centered Design* (RR-03-16). Retrieved December 28, 2005 from <http://www.ets.org/research/researcher/RR-03-16.htm>

Evidence-Centered Design implies that classroom formative assessment and summative assessment should work in a synergistic manner because both types of assessment should derive from the same conceptual base. In other words, as Matters implies, both should be created to reflect a single definition of domain proficiency (in theory as well as in implementation). That does not mean that a summative test can provide evidence to support the same claims as a formative test. A claim of proficiency in a broad domain requires a test that samples comprehensively, if not deeply. In contrast, a claim of proficiency in a narrow segment of that domain – which is what a formative test is typically intended to offer – requires deep sampling of a limited number of proficiency components.

Could the reverse situation, however, hold? Could many formative tests provide a significant portion of the evidence needed for summative decisions? Could a synergistic system of heavier formative assessment and lighter summative assessment satisfy the diverse evidentiary needs of education's multiple information consumers? Certainly, some Australian assessment systems have made impressive progress toward what Matters describes as – and I agree should be – an assessment 'ideal'.

So what additional work needs to be done? First and foremost, we need to identify from the existing research the most promising cognitive-domain models and ascertain how they might articulate with the current content-based conceptions that underlie most government-sanctioned views of proficiency. Understanding that articulation will help in determining how our current curriculum, instruction, and assessment designs may need to change. Where suitable cognitive-domain models do not yet exist, we need to do the research necessary to develop those models.

A second focus for new work might be on methods for efficiently authoring complex online performance tasks like simulations. At present, these tasks are extremely labour-intensive and, thus, expensive to create. We will need to develop authoring tools and generalised simulation environments that permit new tasks to be configured quickly from more primitive task elements.

A third focus should be measurement models for accumulating evidence over time. Is it sensible to assess a student formatively on each new occasion as if we had never assessed him or her before? Couldn't the results of previous assessments add to what we learn from the current one? Students do, however, forget some of the things they have learned, so our models would need to take such factors into account.

A final focus is on the teacher. As others have argued, teacher judgements can be important evidence of proficiency too. We need to develop principled ways of incorporating those judgements into the evidentiary record on which we base our formative and summative inferences about student proficiency. But we also need to support teachers in providing that evidence. A last benefit of grounding our assessments in well-developed cognitive models is that such models can offer to teachers a more powerful way of conceptualising domain proficiency, conceptualising how to assess student standing in it, and conceptualising how to help students achieve it.

With these thoughts in mind, on to Gabrielle Matters' insightful review of using data to support student learning and decision making in education.

Randy Bennett

*Distinguished Presidential Appointee in the Research & Development Division at
Educational Testing Service in Princeton, New Jersey.*

Dr Bennett's work for over two decades has focused on the use of new technology to improve assessment including research on the comparability of computer and paper tests, the use of multimedia and simulation in testing, and the scoring of responses to open-ended tasks automatically.

Contents

	Foreword	iii
Section 1	Introduction	1
	Terminology	1
	Organisational framework	2
	Using the organisational framework	2
	Why data and why now?	3
	Plus ça change, plus c'est la même chose	4
	The structure of this review paper	5
Section 2	Data	6
	Playing with words	6
	Data are what you use to do your thinking	7
	Changes in data use	7
	Student performance data	8
	Other forms of data	9
	Attitudes to statistics	9
	Data are the basis for professional conversation	10
	Using evidence-based models for practice	10
	Data sources	12
	Hypothetical question	13
	Conclusion	14
Section 3	Themes	15
	Clustering of the themes in the research	15
	Positioning of the conference papers	15
	Using the diagram on positioning of conference papers	16
	Schools and the system	17
	Teachers, classrooms, schools	18
	Teachers and students	19
	Conclusion	21

Section 4	Dichotomies	22
	Paradigm	23
	Four apparent dichotomies	25
	Dichotomy 1: Assessment for learning versus assessment of learning	25
	Formative and summative assessment	25
	Summary of, and ways of moving forward with, Dichotomy 1	28
	Dichotomy 2: Student work versus test scores	29
	Defining terms	29
	Low research profile of student work	30
	Summary of, and ways of moving forward with, Dichotomy 2	32
	Dichotomy 3: Progress of an individual student versus success of a school	32
	Defining individual student progress	33
	Defining comparatives and superlatives	35
	Defining the success of a school	35
	The good school	38
	Students and reasons for progress	40
	More about values and valuing	41
	Summary of, and ways of moving forward with, Dichotomy 3	42
	Dichotomy 4: Sociology versus psychology as an explication of success	42
	Terminology	43
	Research context	43
	The exciting dangers of using biological data to support learning	44
	Paradigms of sociology and psychology	44
	Summary of, and ways of moving forward with, Dichotomy 4	45
	Conclusion	46
Section 5	Implications	47
	Methodological perspective	47
	Professional development	48
	Teachers' attitudes and expertise	48
	Summary comments on methodological perspective	49
	Strategic perspective	49
	Policy makers and data	50
	Evidence–practice gaps	51
	Summary comments on strategic perspective	51
	Ethical perspective	52
	Ethical guidelines	52
	Optimism	52
	Summary comments on ethical perspective	53
	Conclusion	53
	Concluding comments	54
	List of 2005 ACER Research Conference papers	56
	References	57

Introduction

This review paper identifies and captures the major themes from the ACER 2005 Research Conference, 'Using data to support learning'. The title of the review, 'Using data to support learning in schools: students, teachers, systems', is an extension of the conference title, the significance of which emerges as this review unfolds. This review distils ideas and information from the conference papers and also from conversations that took place during the conference, and it looks at what are claimed to be effective ways of using data to improve student learning in schools. *How* data are used to support student learning is what is under consideration, from collecting reliable data, making valid interpretations, reflecting on those interpretations, and acting upon that information if and when necessary.

The purpose of Section 1 is to set the stage for the discussion of how data are used to support student learning in schools by reviewing some of the reasons for the increased attention given to data analysis in education today, and by providing an organisational framework to describe the use of data in education. Section 1 concludes with an explanation of what this review includes and what it does not include.

Terminology

Attitudes to data analysis in education have changed in recent years, and it is illuminating to consider the background against which these changes have occurred.

But first, what do we mean by the Latin plural 'data'? In this review paper, the term 'data' is taken to encompass the trio of *input*, *process* and *output*.

Data inputs to education include funding, personnel, teaching resources, facilities, teacher quality and training, and less tangible inputs such as students' intellectual, social and cultural backgrounds.

Data outputs of education include grades, test scores, retention rates, tertiary entrance ranks and student destinations.

Situated between inputs and outputs is a variety of processes during and through which student learning occurs. These processes include the student–teacher interaction, the student–curriculum interaction, and the teacher–curriculum interaction.

The author uses the term 'curriculum' to mean curriculum as content, as well as curriculum as process. By curriculum as content is meant the carefully selected traditions of knowledge and texts, skills and competences, processes and practices that education systems deem to be of value for construction by, and transmission to, successive generations of learners. Curriculum

as process is the totality of the experience that the student has as a result of the provision made. The pre-active (i.e. intended) curriculum is what is planned to happen in national, State/Territory, district, school or classroom arenas. The interactive (i.e. enacted) curriculum is what actually happens.

Attributes of society (taken to be the broader community, culture, and nation or state in which the trio exists, and by which it is influenced) include society's values and expectations, and educational leadership at government, sector/system, region/district and school level.

Organisational framework

The input–process–output model for data in the student learning environment is represented diagrammatically (see Figure 1). The entries in the boxes are illustrative and the lists are not exhaustive.

The framework is an adaptation of the 3P model of learning and teaching developed by John Biggs (Biggs, 1999; Biggs & Moore, 1993), which portrays learning as an interactive system, identifying 'three points of time at which learning-related factors are placed: *presage*, before learning takes place; *process*, during learning; and *product*, the outcome of learning' (Biggs, 1999, p. 18).

Biggs's model draws attention to two sets of presage factors: meta-contextual factors and those factors specific to the learner. In the adaptation of his model to datasets, the presage components are data about students, teachers, and school organisation and resourcing. His model of classroom learning describes a system in equilibrium. There is, using the analogy of a chemical reaction, a back reaction as well as a forward reaction: feedback from product (i.e. learning outcomes) to presage (e.g. teacher quality). This model is capable of generating predictions and of providing feedback, both of which are relevant to the study of student learning. Reading from top to bottom, from input through process to output, the diagram portrays the storyline (in the language of datasets) for an individual student or student cohort.

Using the organisational framework

Two examples follow of how this organisational framework locates the data that are used in educational research. The reason for including the two quotations here is not to analyse their contents as part of our story about the use of data to support learning, but because both of these studies relate in some way to student learning. Both required a collection of facts and figures and observations; that is, they required the collection of data. In each case, the terms that fit the labels in the model are not italicised.

The first example is from a paper on league tables and their limitations by Goldstein and Spiegelhalter (1996):

The OECD also identifies a shift from the use of input indicators such as expenditure, to a concern with outputs such as student achievement. Interestingly, the report is little concerned with process indicators such as curriculum organisation or teaching styles.

(Goldstein & Spiegelhalter, 1996, p. 5)

Information embedded in the first example about expenditure and student achievement would fit into the model as data input and data output, respectively, and curriculum organisation and teaching styles would both fit as student-learning process.

The second example is from Cohen, Raudenbush and Ball's (2003) study of resources, instruction and research:

Many researchers who study the relations between school resources and student achievement have worked from a causal model, which typically is implicit. In this model, some resource or set of resources is the causal variable and student

achievement is the outcome. In a few recent, more nuanced versions, resource effects depend on intervening influences on their use. We argue for a model in which the key causal agents are situated in instruction.

(Goldstein & Spiegelhalter, 1996, p. 5)

Information embedded in the second example about school resources, student achievement and instruction would fit into the model as data input, data output and student-learning process, respectively.

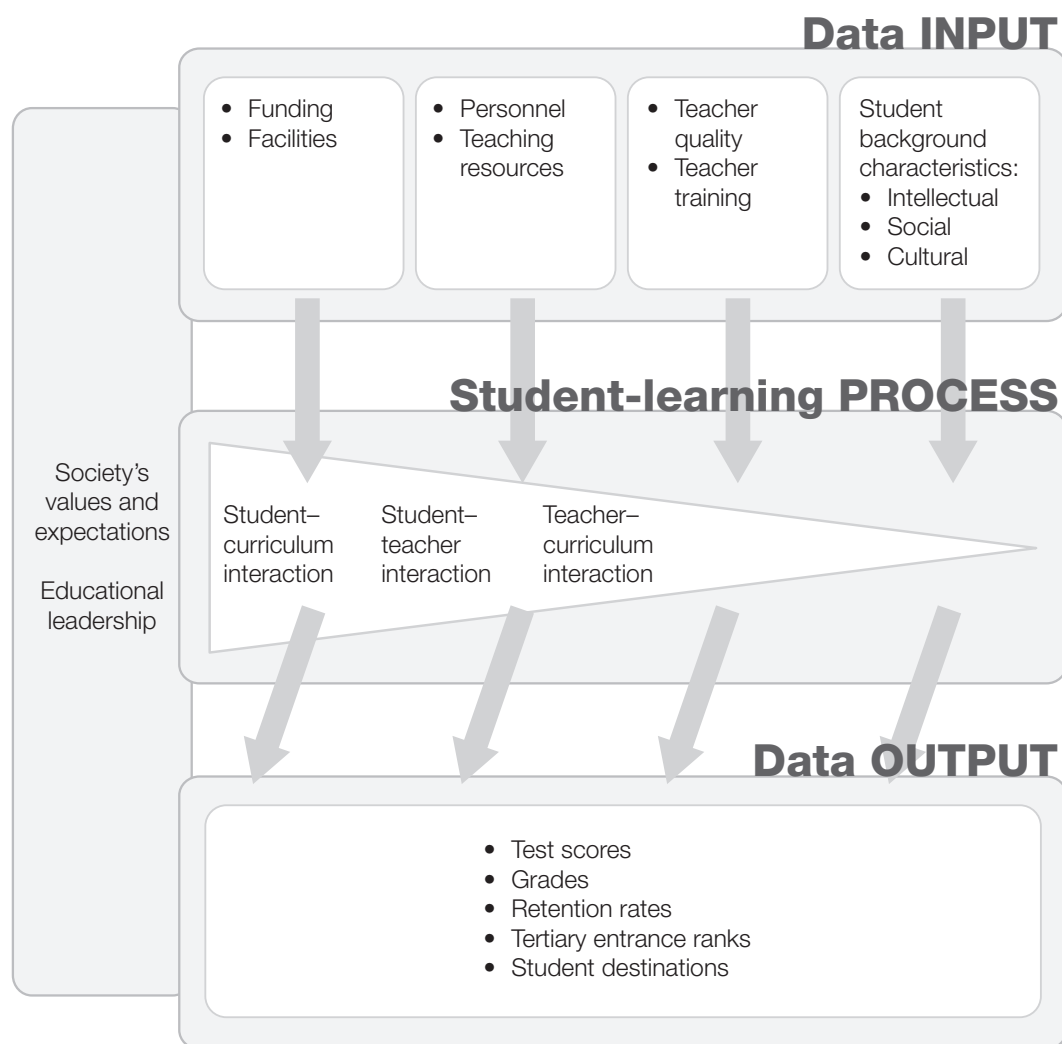


Figure 1. Organisational framework for data in the student-learning environment

Why data and why now?

It is almost impossible to think of the phrase 'using data' in the education context without thinking about assessment. And it has become almost impossible to think about assessment, especially student assessment, without thinking about accountability to a range of parties. In her keynote address at the ACER 2005 Research Conference in Melbourne, Lorna Earl asked the double-barrelled question, 'Why data and why now?'. The professional experience she described resonated with many of the baby boomers in the audience. They too had had 30 years' experience of working with data but only in the past five years had they worked in a society which emphasised the link that has always existed between using data and supporting learning.

It is certainly the case that by 2005, 'schools are awash with data' (Hattie, 2005, p. 11) in an era of 'data as a policy lever' (Earl, 2005a, p. 6).

In the past several decades, a great deal has changed; the 21st century has been dubbed the 'information age'. There has been an exponential increase in data and information, and technology has made it available in raw and unedited forms in a range of media. Like many others in the society, educators are trying to come to grips with this vast deluge of new and unfiltered information, and to find ways to transform this information into knowledge and ultimately into constructive action.

Accountability and data are at the heart of contemporary reform efforts worldwide. Accountability has become the watchword of education, with data holding a central place in the current wave of large-scale reform.

(Earl, 2005a, p. 6)

Thus Earl answers her own question as to why there is such an emphasis on data analysis in education at present. And there is nothing philosophical or existential about the answer. Quite simply, we are living in an age of accountability while simultaneously experiencing a boom in technology. Each feeds off the other. This situation is revisited in Section 2.

Plus ça change, plus c'est la même chose

The observation that 'the more things change, the more they stay the same' is usually applied when an imposed change does not result in an improved situation.

In his introduction to the third edition of *Educational Measurement* (American Council on Education, 1989), Robert Linn listed the three biggest changes in the 18 years since the second edition of *Educational Measurement* (Thorndike, 1971) as:

- attention to Item Response Theory
- computerised test administration
- the fair use of tests with minorities.

He stated:

There are senses in which there has been tremendous change and others in which there has been relatively little.

(Linn, 1989, p. 1)

This review argues that the changes for the period (1971–1989) as noted by Linn (1989), and the changes that have occurred in the period since then (1990–2005) are of a similar kind. That is, there have been a few hugely significant changes, which stand in stark contrast to the many others of varying impact. Two of the 'hugely significant changes' over the past 30 years are identified and discussed here. The first is the tremendous public influence of, and the sophisticated methodological breakthroughs in, psychometrics and educational statistics. The other is the impact of testing and advances in test administration and marking (e.g. through the use of computers). Both of these changes have had a very broad influence on society and education in general and on the use of data to support learning in particular.

We now leave behind the two periods that spanned the past 30 years and focus on two decades, the last decade of the 20th century and the first decade of the 21st century. Linn (1989) put forward the view that 'the biggest and most important single challenge for educational measurement ... is to make measurement do a better job of facilitating learning for all individuals' (Linn, 1989, p. 1). This was the challenge for the 1990s. What was the challenge to be in the first decade in the 21st century?

It could be inferred from the volume of work currently being undertaken within the education field and from conversations at the conference that the challenge for the 'noughties' is not unlike what it was for the 'nineties' – to do a better job of facilitating learning for all individuals. Both focus on the enhancement of student learning. For the 1990s, however, the

agent of enhancement was deemed to be the practice of measuring performance, whereas, for the early 2000s, the agent of enhancement is deemed to be the practice of using the collected data, especially data about student performance.

The two practices – measuring performance and using data – are not the same. In terms of the input–process–output model (Figure 1), the former practice predominates in the generation of information about student learning and predictions based on the analyses; the latter practice predominates when reflecting on the information generated and feeding it into the learning process.

The change in emphasis described above represents a subtle shift from assigning major responsibility to those who do the design and apply the measurement models to those who receive and act upon the products of the measurement models. Nevertheless, at least two other things have remained unchanged over the intervening decade or so. First, student achievement was and is at the heart of all educational expectations. Second, it has been and remains important for educators to understand how assessment data (i.e. data about student achievement) can be used productively.

Regardless of how much things have changed or remained the same, regardless of the number of explanations about learning that continue to elude us, this review is based on the premise that using data can be a creative and imaginative process.

The structure of this review paper

The main purpose of the review is to enable readers to stop and consider both the conference papers and the bigger issues about using data, which most of the conference papers addressed to some extent.

First, some conceptual and historic context needs to be developed, and the author uses an extensive reading of the literature and examples of practice to provide that context. Then, according to their main focus, she situates the conference papers on a map of the education delivery system that she had previously created, according to their main thrust. The final section of the review paper questions how those issues can be addressed in the future.

This review accommodates a range of perspectives on data, which is intended to appeal to a range of readers. The nature of this wide and varied audience – classroom practitioners, whole-school stakeholders, parents, and systemic bureaucrats – influenced what was collected, compiled and included in this review. In preparing this review, the author has been mindful of not overstating the potential of such a document to capture the full sweep of the issues. Length constraints prevent it becoming a primer on collecting and interrogating data or a textbook on the factors that influence student learning, or a thesis on the link between using data and improving learning. The range of references that exist on these topics is vast. The author has been selective in her coverage and the relatively short recounting of some of the issues in this review has, to a certain extent, been coloured by her broader brief to go beyond the research and the conference, into the realms of her personal experience.

As a result of the high degree of interrelatedness of the concepts of contextualising data, data sources and the use of data to support learning effectively, it is not easy to treat particular facets of the review in isolation. To a certain extent, each of the five sections in this review paper shapes the other.

Section 1 has set the context, provided definitions and an organisational framework. It was a stage-setting exercise and framed what will be used in subsequent sections of the review. Section 2 focuses on some purposes for analysing educational data, the role of data in professional work, and potential data sources. Section 3 records what the research says about how data can be used to support learning and notes recurring themes. Section 4 identifies and comments on some of the apparent dichotomies in our current discussions about data and evidence and suggests some alternatives for operating more effectively. In conclusion, Section 5 sets out some implications for policy and practice of the methodological, strategic and ethical aspects of the issues discussed in earlier sections.

Data

This section builds on Section 1 in a discussion of the purposes of analysing data in and about schools, by reviewing the role of data in professional work, and by applying a list of potential data sources to a hypothetical question. It also identifies and describes key elements in evidence-based decision making.

Playing with words

The title of this review, ‘Using data to support learning in schools: students, teachers, systems’, as well as being an extension of the conference title, is deliberately ambiguous.

However, for many people, on first reading, the review title immediately and unambiguously conjures up the image of some particular combination of user/subject/source/agency: teachers using data, data about students, data that come from tests, data that support the learning of students in schools. This review strives to have readers go beyond that image to include, for example, systems using data, data about systems, data that come from student performance measures other than test scores, plus data about teacher practice and system behaviour. A consideration of the component parts of the title follows.

Using data: Who is using the data and why? The teacher, the sociologist, and the policy analyst, amongst others, are all users of data, although their reasons for doing so may not necessarily be the same.

For a teacher, the central purpose of analysing data is to improve the learning of one or more particular students; that is, the individual teacher and the school take the students who come to them and seek to improve the learning of those students. This purpose is different from that of the sociologist seeking to understand patterns of participation, or that of the policy analyst seeking to understand the impact, if any, of policy setting. A social scientist wants to understand patterns of participation. A policy analyst wants to know the impact of some policy settings.

(Allen, 2005, p. 87)

It is possible, of course, for one person to take on all three roles (teacher, sociologist and policy analyst). And it is obvious that there are other reasons – like pure intellectual curiosity

– for using data in schools. But regrettably, those other reasons are not so prominent in today's discourse.

To support learning in schools: Whose learning is being supported here? Although the obvious answer is students, it could well be teachers, principals, parents, those who run school systems, and anybody else who needs or wants to be knowledgeable or to support learning.

Data about what or whom?: The prime subject of data gathering would be students and student learning. It could also be teachers, teaching strategies, principals, parents, systems, money, school buildings, school communities, to name but a few. What are the sources of these data? Tests, questionnaires, observations – the list is endless. A large number of potential data sources are listed later in Section 2.

And, following the colon in the title, 'Using data to support learning in schools', the preposition has been deliberately omitted. Is the preposition to be understood as 'by' or 'of'? It could be either, depending on whether we are thinking about the users of the data analysis or the subjects of the inquiry. It could also be both. The important point is that student learning might be improved if people (educators and non-educators alike) used data as the basis of their decision making as opposed to mere opinion. Barry McGaw made this point at the ACER 2002 Research Conference when he said, 'Without data, I'm just another person with an opinion'.

The kinds of data to be collected and interrogated can be data that emanate from different sources:

- observing student performance (which is the outward and visible sign of student learning)
- research into factors that improve student achievement (such as teaching practices and student motivation)
- research into factors that affect participation rates (such as gender and socioeconomic status)
- evaluation of government policies (such as school reform, curriculum revision and testing regimes).

Data are what you use to do your thinking

Data shape the landscape of our professional lives. Or if they don't, they should, given that education is a profession. Sometimes, like geological upheavals in a landscape, data surprise us and mark *our current views* as falsely secure and prone to reversal. At other times, like sedimentary rock, they mark our views as static and unchallenged. Sometimes, data provide evidence that the *learning process* itself is uneven and able to be re-directed. At other times, they provide evidence that the learning process is pre-wired and unable to be influenced or re-directed.

Datasets form part of our system of decision making, and we must be prepared to grapple with the concepts of validity and reliability, correlation and causation in order to understand students' needs, to provide information to parents, systems and policy makers, and to promote better teaching, or teaching with a different focus. Should the data reveal that students do not seem to know what they are expected to know, understand it, or use it, we must examine our teaching strategies, curriculum design, and possibly our expectations of students.

Datasets also form part of our armoury of meaning-making, and we must be prepared to investigate the things that continue to elude us, such as understanding the differential capacity of students to organise knowledge, in order to appreciate fully the very nature of knowledge and learning.

Changes in data use

In the not-too-distant past, educational data were slow to turn around, unwieldy to manage, and too disparate to enable meaningful comparisons to be made between groups or over time.

Today, owing to advances in computing and communications technology, the widespread use of data for decision making is possible and practicable at every level of the education system – from students and teachers to parents and school administrators to stakeholders and policy makers.

Furthermore, there is a new generation of computer software and Web-based products designed to integrate data into the core elements of teaching and learning. These products are being promoted as capable of linking curriculum intent to classroom practice and learning outcomes. There are computer packages that give instant access to the world of statistics and mathematical modelling. These products allow a recipe-book approach to data collection and analysis which, if used wisely, delivers information to people otherwise locked out, but which, if used unwisely, delivers false confidence and shallow understanding. It does little good to be able to calculate a statistic if its meaning cannot be correctly understood and communicated to others.

Student performance data

Today's taxpayers and governments are demanding accountability of their schools; they are looking to numbers (they ask 'What do the stats say?') for evidence that things work, that money has been well spent, that learning outcomes have improved in nature and depth.

Savvy and well-informed educators are embracing performance data as a useful means for directing school improvement. The ability to track individual student performance, aggregate and disaggregate data more easily, and use sophisticated and high-speed data-collection systems present a new host of options for using and interpreting data.

Now that such information is available, there is no going back to decision-making styles that rely strictly on gut feelings or anecdotal information.

(Killian & Bellamy, 2000, p. 12)

Comparative data across schools and jurisdictions make it easier to discern the practices and policies that work from those that do not, and to speculate about why it might be so. Longitudinal studies make it possible to appraise the nature and degree of developmental changes in relation to the expected course for the practice or policy under scrutiny.

An interesting if not distressing by-product of the latest wave of comparative data such as data from the OECD Programme for International Student Assessment (PISA) (ACER, 2005) is the quantification of failure, which seems to generate performance anxiety at country, state, system, school, and teacher level, but which is possibly most keenly felt between schools. Eisner (2001) is also inclined to question the 'blinkered vision of school quality that now gets front-page coverage in our newspapers'.

Perhaps our society needs losers so it can have winners ... I believe that those of us who wish to exercise leadership in education must do more than simply accept the inadequate criteria that are now used to determine how well our schools are doing.

(Eisner, 2001, p. 372)

This is not to say that the judicious use of PISA data is not illuminating. McGaw's (2005) checklist is pertinent here, and it is:

- Why are you looking at the data?
- Who is looking at the data?
- What in particular are you looking at?

Other forms of data

Student performance data are not the only data whose analysis has the potential to support learning in schools. Ken Rowe (2005) made a point of referring to the importance of collecting and analysing data about teachers as well as data about students, while Gabrielle Matters (2005) noted how good research data gathered during a reform process can lead to good policy making in the service of improving standards. Lingard, Luke and Ladwig, in one of the largest observational studies carried out in Australia – the Queensland School Reform Longitudinal Study (The University of Queensland, 2001) – investigated the degree to which reform of central office support and school organisational capacity is capable of generating pedagogical change and improved student outcomes. Lawrence Ingvarson (2005) studied the application of principles of effective professional learning to reform strategies that emphasise professional capacity building among principals and teachers.

The preceding examples, chosen to illustrate how data can ‘fuel the reform process’ (Killion & Bellamy, 2000), also highlight the usefulness of data other than student performance data.

Attitudes to statistics

Psychologically speaking, it is interesting to note a diminution over the past 20 years in the fear of statistics – which gave rise to book titles such as *Statistics without Tears* (Rowntree, 1981) – and in the mistrust of statistics – which gave rise to cartoons such as the one about a statistician drowning in water of average depth x while in the act of using a ruler to measure the water’s depth. Parallel to these reduced levels of fear and loathing is the realisation that statistics do not provide knowledge; they only help make sense of observations and specify the kinds of conclusions that can be drawn from observations. Also, in our time-poor society, it is convenient to be able to take a large amount of data and, with the push of a button, reduce it to a few numbers, graphs, or tables that are interpretable. Cynics amongst us might explain away this change in attitude as an exercise in pragmatism.

Interestingly, this new attitude to the outputs of data analysis has occurred in spite of a retreat from the quantitative world. One commonly held view is that some people were seduced by numbers and became overly reliant on quantitative solutions. A contrary view is that some people just did not have the ability or willingness to understand numbers and rejected quantitative solutions out of hand.

Whatever the explanation, it is inarguable that there presently exists a strong demand, from policy makers and practitioners alike, to know what works (which can translate into causal inference), to know when it works and for whom (which can translate into causal generalisation), and to know how it works and why it works (which requires other methodologies).

Others express a view about the links between greater familiarity with data and its increased influence on reform.

Fear and mistrust of data are giving way to a new culture of [data] use ... to achieve goals.

(Killion & Bellamy, 2000, p. 12)

The appearance of this new culture is a relief because statistical data on school programs and student performance provide educators and governments with their only real evidence of the health of the education system and the success or failure of educational programs. The skilled use of data is important because caution must be exercised when making pronouncements about the performance of schools and educational programs. Moreover, there is pressure on individuals and systems to provide only good news to their various publics (e.g. parents and governments). So, even if the data reveal a poor story, the pressure for good news can lead to the utterance of motherhood statements (‘there are good things happening out there’) or, even worse, public statements that contradict the evidence.

Michele Bruniges (2005) grappled with the nature of evidence in terms of the purpose for which it is intended. She was unequivocal in her answer to the question, 'Data about what or about whom?' For her, it should be information about students as interpreted by teachers.

A Greek philosopher might suggest that evidence is what is observed, rational and logical; a fundamentalist – what you know is true; a post-modernist – what you experience; a lawyer – material which tends to prove or disprove the existence of a fact and that is admissible in court; a clinical scientist – information obtained from observations and/or experiments; and a teacher – what they see and hear.

(Bruniges, 2005, p. 6)

Data are the basis for professional conversation

Professionals aspire to practice supported by research; that is, evidence-based practice¹. The term 'evidence-based practice' began its life in healthcare as evidence-based medicine. Evidence-based medicine shifts decisions about healthcare practices away from opinion and past practices to practices based on evidence. Examination of the model is useful for analysing the use of data (to support learning) in education.

According to John Hattie (2005), the notion of evidence-based decision making is 'a current fad radiating out from the United States'. To those like Hattie who cannot see what all the fuss is about because using evidence has always come naturally to them, it might well be the case that evidence-based anything is a fad. It is indisputable that the term 'evidence-based practice' is an Americanism. A Google (Web) search on the topic yielded about 8,900,000 results, 99.9 per cent of which were based in the United States of America, including the research of Antman, Lau, Kupelnick, Mosteller and Chalmers (1992), which showed that physicians in the United States of America failed to recommend medications up to 10 years after they have been shown to be efficacious and continue to recommend treatments up to 10 years after they have been shown to be useless.

Whether or not research would yield the same results about Australian physicians is not the message to be taken from the work of Antman et al. (1992). The interesting question is whether or not research would yield similar results about evidence-based practice in education, here in Australia or elsewhere.

Using evidence-based models for practice

Evidence-based medicine (EBM) integrates clinical experience and patient values with the best available research information. The EBM process has four steps for incorporating the best available research evidence in decision making:

- 1 Asking answerable questions*
- 2 Accessing the best information*
- 3 Appraising the information for validity and relevance*
- 4 Applying the information to patient care.*

(Craig, Irwin, & Stockler, 2001, p. 248)

Once the term 'patient care' is replaced by 'student learning', this formula for using research evidence in decision making in medicine translates across to education. Section 5 in this review takes up the issue of evidence–practice gaps in education in Australia.

¹ Now with various designations such as 'research-informed policy' and 'evidence-centred design'.

In his opening address at the conference, the CEO of ACER, Geoff Masters (2005), made the general point that *all* genuinely professional work depends on the collection, analysis and use of relevant information about [the]:

- situation being confronted
- effective strategies for dealing with the situation
- progress of a course of action
- outcome of that course of action.

These are the same four steps that define EBM, and Masters (2005) argued the value of making connections to that work in his conference presentation. It is therefore useful to describe the four stages in the process in some detail.

First, the *situation*: To find out what they are dealing with, professionals employ specialised tools and techniques to collect data and then they apply their expertise to reading and interpreting those data – turning data into information. We should be mindful of the following stricture:

Data are not information; information is that which results from the interpretation of data.

(Mitroff & Sagasti, 1973, p. 123)

Figure 2 relates to the use of a specialised tool (X-rays), by a professional (in this case, a radiologist, a medical doctor who specialises in the field of X-rays), to collect data (in this case, three images on a photographic film), to which she or he will apply their expertise to read and interpret the images, thus producing information about the bones in the patient's hand, shoulder and pelvis – and sometimes about related conditions.

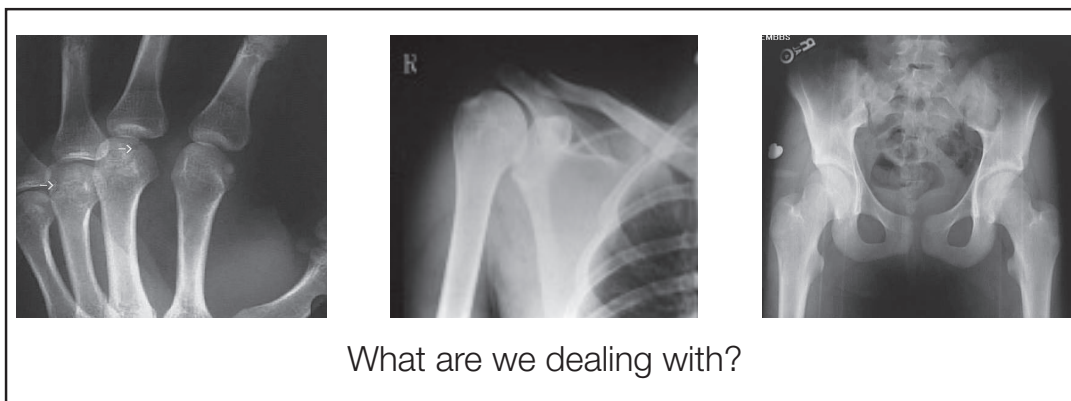


Figure 2. Medical example of the first stage in evidence-based decision making

In the process of interpreting data, there is a blending of facts and theoretical rationale.

Facts do not speak for themselves; nevertheless, facts must be given a hearing, or the scientific point of the process of interpretation is lost.

(Kaplan, 1964, p. 385)

The result of this processing is evidence. Because evidence is produced in answer to questions (what if? how many? which type?), the way we pose the question reflects our values as well as our intent and thus helps to determine the answer obtained. Kaplan (1964) acknowledges this perennial problem:

[Through] a process of interpretation, data have meaning, and this word 'meaning' like its cognates 'significance' and 'import', includes a reference to values.

(Kaplan, 1964, p. 385)

Second, the *strategies*: Professionals develop solutions to fit specific problems rather than routinely implement one-size-fits-all solutions. To decide which treatments/interventions/practices are needed on the basis of the information gained, professionals draw on their professional knowledge base as described earlier in this section.

Third, the *progress*: Professionals seek evidence about the progress of a treatment/intervention/practice and, on the basis of this information, might consider alternative approaches, study what has been effective, and under what conditions, in the past. The monitoring of progress often requires measures that can be compared over time.

Fourth, the *outcome*: Professionals require information about the outcome of a treatment/intervention/solution in order to evaluate how well it has worked. What did we learn? Was the initiative effective? Was it more effective than the alternatives? Information about how well an initiative has worked adds to our professional knowledge base.

There is a close correspondence between the stages of evidence-based decision making (situation, strategies, progress and outcome) for professionals in general as defined above and the processes of producing evidence about education.

Data sources

A wide variety of data can be collected about learning, learning experiences, and the contexts in which learning takes place. Where can we find the data? Lorna Earl (2005a) provided the conference with a PowerPoint slide that listed a multitude of data sources, which she classified under nine headings (see Table 1).

Readers can readily insert their own additional entries. The author has inserted three: 'Student work' in the category 'Student achievement'; 'Itembanks' under 'Teaching and assessment practices'; and 'Professional development' under 'School culture'. Her additional entries are italicised.

The purpose of Table 1 is to articulate explicit links between sources of data (as in Earl's classification) and locations of datasets (as in the input–process–output model for data in the student-learning environment). To this end, the now 43 potential sources of data (entries in the middle column in Table 1) were clustered according to the three elements of the input–process–output model (Figure 1 in Section 1). As a result of this analysis, the entries in the right-hand column of Table 1 reflect the destination of their cluster in the 3P diagram. It is not claimed that the resulting matrix represents a unique clustering. (One of the data sources, 'beliefs about teaching and learning', is assigned to the appropriate part of the trio of Ps as the model would require; it is also assigned to the societal milieu in which the trio exists.)

The value of the resulting table is that it can act as an aide-mémoire when deciding what data should be collected after questions have been framed about student learning.

Table 1. Reconciliation of data sources and data locales

Data source	Potential source of data	Data point
1 Student demographics	Attendance Enrolment Grade level Ethnicity Gender First language Health issues Socioeconomic status	Input Input Input Input Input Input Input Input
2 Student achievement	Standardised, norm-referenced, criterion-referenced tests Questioning in class Performance and standards-based assessments Teacher-made tests, projects, quizzes Teachers' observations Grades and grade-point-averages <i>Student work</i>	Output Output Output Output Output Output Output
3 Teaching and assessment practices	Instructional and learning strategies Instructional time and environment Organisation of instructional components Assessment practices Classroom management philosophies <i>Itembanks</i>	Student-learning process Student-learning process Student-learning process Student-learning process Student-learning process Student-learning process
4 Parent opinions and behaviours	Parent perceptions Parent involvement in the school Parent support of student learning	Output Student-learning process Student-learning process
5 School culture	Relationship between educators Relationship between students and educators Beliefs about learning and teaching <i>Professional development</i>	Student-learning process Student-learning process Student-learning process; Societal milieu Student-learning process
6 Staff demographics	Background Interests Qualifications Gender Ethnicity	Input Input Input Input Input
7 Programs	Program descriptions Course outlines Special programs	Student-learning process Student-learning process Student-learning process
8 Resources and materials	Computers Textbooks Software Workbooks Art supplies Musical instruments	Input Input Input Input Input Input
9 Physical plant	Configuration of space Playground	Input Input

Note: Based on Earl (2005a)

Hypothetical question

In order to examine the strength of the aide-mémoire as a useful model for teacher-researchers, we can use Table 1 to study a hypothetical educational research question and decide on a data source.

The selected research question is this: Has student learning improved as a result of the intervention? For this hypothetical exercise, it is not absolutely necessary for the intervention to have a named topic but it is imperative that we consider how to measure student learning and how to measure changes (improvements) in student learning. We decide that we want information about the products of learning so we need to look to a source of data about student achievement/performance. We need to do this because, as yet, we are not in the position

to attach electrodes to the cerebral cortex of a student in order to obtain direct evidence of learning as it is happening.

According to Table 1, we have at least six potential sources of data in the student achievement category:

- standardised, norm-referenced, criterion-referenced tests
- questioning in class
- performance and standards-based assessments
- teacher-made tests, projects, quizzes
- teachers' observations
- student work.

Data about what actually happens in schools can be classified as primary data (e.g. work of students and classroom observations of teachers). Secondary data (e.g. students' test scores and responses from parent surveys) are one step removed from the teaching–learning site. In this hypothetical exercise, we decide to consider using both (a) test scores and (b) student work.

The steps undertaken in this decision-making process – from being given a research question to deciding on what data to collect – are summarised below as a series of questions and answers in Table 2.

Table 2. Steps in deciding possible data sources

What is the research question?	Has the intervention improved student learning?
What do we need to measure?	The quality of student learning (i.e. the success of the intervention)
What gives us the evidence?	Data on the products of student learning
What are our potential data sources?	Entries under the heading 'Student achievement' (e.g. teachers' classroom observations, tests, assessment tasks)
Which data source will we use?	a) Tests and b) Student work
What instrument will we use to get these data (thence this evidence)?	a) Findings from an IEA study in the appropriate field and b) Rich tasks

Conclusion

The purpose of this hypothetical exercise was to model how these questions and the process of collecting and examining data is inherent in all research questions. So the medical model, with few revisions, works perfectly well for the educational professional. The Earl (2005a) and Matters (2005) categories also remind us of the importance of considering all sources of data. If the data collection is narrow, the interpretations available from the analysis will also be narrow. As this section has emphasised several times, the questions need to be both open and focused for, above all, they determine the parameters for the investigation: the data source, what data are collected and the interpretations available to practitioners, policy makers and stakeholders.

This hypothetical exercise limited itself to describing a decision-making process that culminated in answering the question: What data will we use? The list of questions and answers above actually goes one step further into the issue of what instrument(s) to use. And it consciously ignored the fact that we need to embark on a comprehensive literature review to identify factors that need to be incorporated into our research design. Only then would we be able to design methods for collecting data, analysing the data, and drawing conclusions until, eventually, we had the answer to our research question.

Section 3 will consider these other issues in greater detail.



Themes

The focus of Sections 1 and 2 was on contextualising data and data sources. In Section 3, attention is primarily on the papers from the ACER 2005 Research Conference, and thus the Australian context of these papers, bearing in mind the nationality of most of the conference participants. Section 4 will draw on work from the international scene. As previously noted, there is a high degree of interrelatedness between, and overlap in, data contexts, data sources and data use. Nevertheless, Section 3 is an attempt to isolate and then record what the research says about how data can be used to support learning in schools.

Clustering of the themes in the research

This review adopts a broad interpretation of data to include evidence about students and learning, teachers and teaching, and systems and the ‘games’ they play. Against this background, a survey of what the national and international research says about the use of data to support learning in schools reveals that there are recurring themes – about students, about teachers and classrooms, about schools, and about systems/sectors, all relating to how data are used for monitoring, policy formulation, target setting, evaluating and reforming. These themes provide a skeleton for mapping the conference papers, as in Figure 3 later.

Positioning of the conference papers

Section 2 highlighted the ambiguity inherent in the title of this paper – learning for whom and about whom? The conference papers covered many of the permutations arising from this ambiguity. The ‘whom’ in both cases can be at any level in the education delivery system. That is, who is learnt about and who does the learning can be at the student level, the teacher or classroom level, the school level, the system/sector level, or the system owner level.

The diagram that follows (Figure 3) identifies parts of the education delivery system on which specific strategies are, according to the research literature, most evidently focused. And it positions the conference papers (three keynote papers and 18 concurrent papers) in terms of their advocated use of data (learning by whom and about whom) and the level of that use (student, teacher/classroom, school, system/sector, or their interstices). This positioning may seem idiosyncratic in places, but it represents the author’s interpretation of the main thrust of each paper. A list of conference papers by topic and author appears at the beginning of the References.

Using the diagram on positioning of conference papers

The diagram has four essential components. They are listed below and then the significance of each is explained.

- Boxes representing the ‘players’
- Bands representing where the research spotlight shines
- Vertical lines:
 - Toned arrowed, representing accountability
 - Solid, representing coverage
 - Toned unarrowed, representing flow or relationships
- Pointers representing the positioning of the conference papers

Boxes representing the players

The labels on the five rectangular boxes in the central vertical panel refer to the players in the education delivery system – student, teacher/classroom, school, system/sector, or system owner. In terms of the use of data to support learning in the education delivery system, they are the ‘whom’ as in learning ‘about whom’ and ‘for whom’. As previously noted, the ‘whom’ can be at the student level, teacher/classroom level, school level, system/sector level, or system owner level. For example, the teacher/classroom box signals that teachers/classrooms can be data sources. We can use data to learn about teachers/classrooms. Also, teachers (though not classrooms) can be data destinations. Teachers can have their learning supported by using data (from any source) – this is then data for teachers.

Bands representing where the research spotlight shines

The differently shaded horizontal bands refer to the perspectives taken by researchers – what the research puts in the spotlight. Where the use of data is at the level of one only of student, teacher, school, or system/sector, the boxes in the vertical panel would suffice as referents for the focus of the conference papers. But the main thrust of many of the conference papers was directed at the level of an interstice. The labels on the three bands refer to these interstices.

Vertical lines (toned, with arrowheads)

These indicate an accountability relationship, from a player who is accountable to a player who is reported to. Sometimes this is a two-way street (e.g. between school and system/sector); hence the arrows going in two directions. The arrows in the diagram connect different pairs of players; but their different lengths do not reflect any difference in importance.

Vertical lines (solid)

These indicate coverage of the various levels of the education delivery system.

Vertical lines (toned, without arrowheads)

These lines (appearing only in the bottom band) indicate the flow of data from data source (data about whom) to data destination (data for somebody to use).

Pointers to represent positioning of the conference papers

These represent the review author’s positioning of the conference papers. Each identifies, by label, a compressed version of the title or main thrust of a particular conference paper, together with the author’s name in brackets. The other end of each pointer, the source end, is located on either a player box or on a vertical line between/through boxes. No significance should be attached to the fact that some of the pointers are to the right and some to the left of the central panel.

Education Delivery System

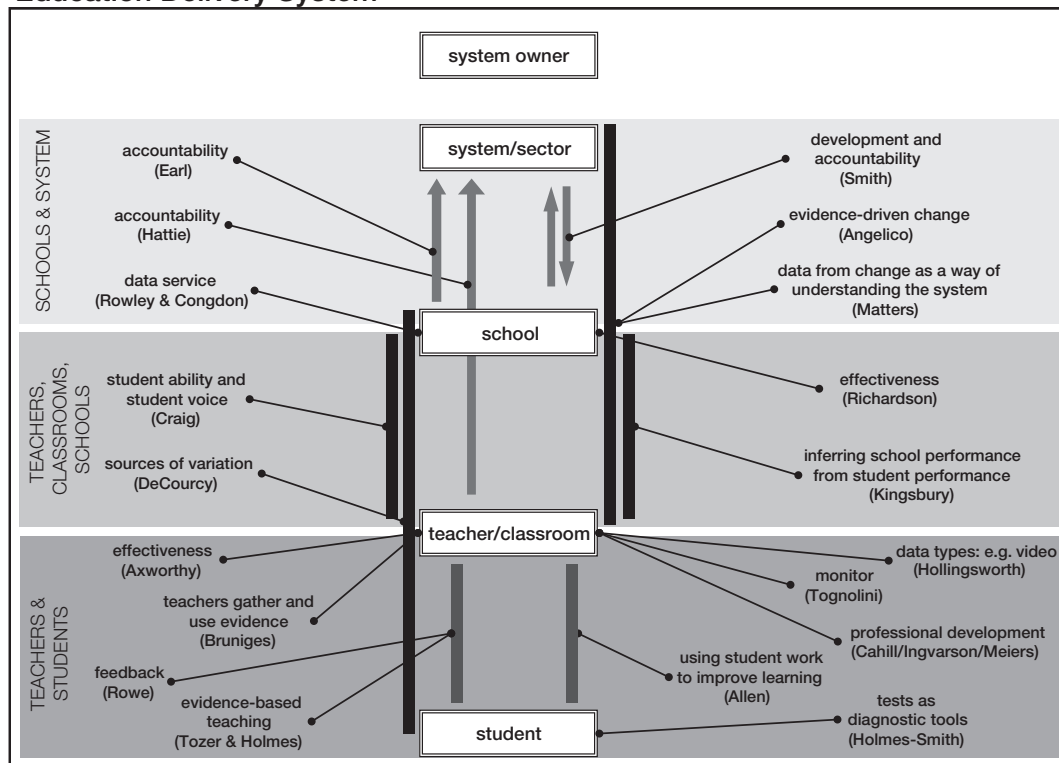


Figure 3. Positioning of conference papers in terms of the advocated use of data

For no other reason than that Figure 3 is drawn that way, the discussion will now cover the main ideas of each of the papers, in a top-down manner with respect to the diagram's layout. But first it is necessary to clarify the term 'system owner' – the personification of the entity responsible for education delivery.

The education delivery system owner ('system owner' in Figure 3) is the government of the day, state or federal, and the current education minister within that government. More broadly, the system owner is the people within the jurisdiction who are represented by that government. The government and the person in the street use data about schools and students in order to learn about schools and students. At an individual level, people learn about their child as a student through the formal and periodic reporting mechanisms, and through the informal, continual interaction in the family. People also learn about their child's teachers in much the same way. Schools use data, often selectively, to help parents and the local community to learn about the school. It may be the school billboard announcing the triumph of the debating team, the newsletter distributed to parents and the local community, or as in the North American bumper sticker announcing that 'My son/daughter is an honour roll student at ...'. People also find out about schools through the publication of 'league tables', whether officially released versions or locally reconstructed versions.

The entity, system owner, attracted little attention at the conference and is not referred to again.

Schools and the system

As Figure 3 shows, the main thrust of six conference papers was the school or the system/sector or the intersection of school and system/sector. Hattie (2005) and Earl (2005a) covered accountability in slightly different ways, but both were mainly concerned about the accountability from the school to the system. Earl made the useful distinction between accounting as a process of gathering, organising and reporting information, and accountability as a conversation about meaning in a broad context. This distinction was echoed by Hattie's emphasis that this is about interpretations rather than data. Earl talked about ways to make data an agent of and

for change, centred on the conversations that arise in the process of accountability, using the metaphor of a painting constructed by the various stakeholders. Hattie argued for models of accountability located at the school and system level to understand questions of effectiveness (what works) and efficiency (what works best). His argument for accountability located at both classroom and school level centred on understanding and interpreting the data that schools are 'awash with'.

Max Smith (2005) echoed the accountability theme and added development as another element in his description of the measurement and reporting package in New South Wales. His paper, however, speaks of an as yet unfulfilled promise – which is that schools are 'presented' with data and are well 'positioned' to use that data as information to support learning. Continuing the theme of supplying data and analytical tools to schools, Glenn Rowley and Peter Congdon (2005) described the data service in Victoria for providing schools with information on students in the senior years (Years 11 and 12). But the question remains, do schools generally, as distinct from a few showcase examples, use these data, learn from them, and consequently change their practice? Again, one yearns for examples not of what is supplied but of what is used, and what value that has.

From another perspective, Matters' (2005) paper discussed the use of data from a change exercise as a way of understanding teachers, schools and the system. She described the learnings gained from the data-rich research project that accompanied the trial of the New Basics. Although not arguing for implementing the New Basics in more schools, all schools or no schools, she described how change can occur (e.g. in the nature and depth of student performance and in the classroom practice of teachers). She described the ways in which change can be accepted or rejected, and she drew on the research evidence for methods that are successful (and methods that are unsuccessful) in bringing about change.

Teresa Angelico (2005) gave a description of the use of research evidence to enhance literacy and numeracy programs and teacher professional development, and also to effect improvement at a school and sectoral level.

Teachers, classrooms, schools

Four conference papers in Figure 3 reference teachers, classrooms and schools. Carmel Richardson (2005) described ability-adjusted means for understanding the effect of school on student achievement. Ability must be inferred from measures other than achievement, and one criticism might be that this is simply taking two measures of the same construct. Her interest, however, is not so much in prediction or association but in variation and explanation, particularly at the within-school level.

John DeCourcy (2005) also looked at sources of variation. He described a set of analyses and graphical displays of student achievement data that can act as an information tool for principals and teachers to enhance pedagogy and, through that, student achievement. Again, one cries out for evidence that this can actually happen on a broad scale, not just in isolated and dedicated instances. No doubt DeCourcy would wish for the same thing.

DeCourcy and Richardson differ on the issue of the influence of the teacher. For DeCourcy 'it's teachers who make the difference'; for Richardson 'claims [of] the proportion of variance explained at the ... teacher or class/teacher level were not supported'. Is it one or the other, or does it depend on the system, since Richardson's analysis was on South Australia data and DeCourcy's on New South Wales data? We know from analyses of international data, which Richardson (2003) has previously coordinated, that the proportion of variance attributable to different levels in the schooling system is, amongst other things, a function of the structure of that schooling system.

Wayne Craig (2005) described a school-level, student-sourced approach to data use that informed teachers and the school. It mapped student progress by using student ability data at entry to the school as baseline data. It also used student opinion data as a feedback mechanism

for teachers individually and the school's performance management process generally. Although it was a case study of just one school, it did offer the 'existence proof' of what can be done.

Gage Kingsbury (2005) also used the 'voice of the student' to infer school performance from student performance in the context of concern about 'failing schools'. Rather than an absolute performance benchmark of a certain proportion of students being above that mark, he argues for a relative benchmark based on student growth. A school 'succeeds' when its students grow, regardless of where they started. A more sophisticated process than merely looking at the percentage of students who are at a certain standard as measured by a single test, this has merit. Some might ask whether we are supporting learning when this happens – and, by so asking, again highlight the ambiguity inherent in the title of this paper – learning for whom and about whom? The use of sophisticated measures of school success can only enhance the learning by systems, schools and teachers about their students.

Teachers and students

In Figure 3, the main thrust of nine conference papers is mapped at the level of teacher or student or intersection of teacher and student.

At the student level, Philip Holmes-Smith (2005) described ways of using tests as diagnostic tools. He discussed how item-level analyses of responses to the standardised (within-State) literacy and numeracy tests administered in each State and reported against national benchmarks can aid schools in understanding their students. He particularly focused on what the 'wrong' answers can tell the school about students' understandings (or misconceptions). The idea of gleaning information about where a student is on the learning journey from her incorrect responses (and the reasons for the responses being incorrect) as well as from what she gets right, is not new. What is new is its application to add value to the ubiquitous basic skills testing regime.

David Axworthy (2005) also discussed the use of standardised tests in literacy and numeracy at the teacher, school and regional level to assess the effectiveness of programs in delivering learning across the curriculum. His analysis is predicated on the view that these tests have breadth across the curriculum in addition to depth in the two areas of literacy and numeracy. He acknowledges that the psychometric qualities of these tests are not in themselves sufficient to convince teachers to use the test data. There needs to be a stage of turning data into information driven by what the teachers want to know about their students, rather than what the system wants to tell them. Systems and analysts need to convey to teachers information that they can use to support their students' learning.

Michele Bruniges (2005) argued for teachers both as a data source or, in her terms, evidence source, as well as being those best placed to interpret and understand other evidence. Hers is an argument for the centrality of the teacher in the making of evidence-based judgements about teaching and learning. One might respond that, yes, good teachers make a difference but what about the rest? At its heart we should take from Bruniges the realisation that teachers must always be participants in evidence-based judgements, not only to capture the particular knowledge and perspective that they bring but also to achieve their 'buy in' to the outcome. Her paper stands out from the others as being primarily principle-based rather than data-based (though she is talking about assessment data).

Rowe (2005) presented an extensive review of the types of data that can be available to teachers, and echoed Bruniges' point that teachers (and schools) must be participants in the data feedback process in order to feel ownership and control. It is then, he argues, that change will penetrate the classroom door, as distinct from bureaucratic fiat, which rarely do. And it is then that data will support teachers in supporting their students' learning.

Lynn Tozer and Marilyn Holmes (2005) provided an example of the teacher acting as the gatherer of numeracy data, and how these data could be aggregated at a class, school or national level. (This is possible in New Zealand but more problematic in Australia.) While they did note that 'good data ... help teachers make good decisions', Tozer and Holmes did not deal

with the issues of the reliability of teachers as raters and the validity of teachers' judgements. Reliability and validity, while not attractive concepts for teachers as Axworthy (2005) pointed out, are nevertheless crucial prerequisites for data-driven judgements and decisions. As with poor data, poor judgements do not support learning.

Jim Tognolini (2005) presented an example of an online assessment tool that could also be used by teachers to monitor students' progress and thus provide a feedback mechanism for teachers. As a preliminary exercise, he presented a conceptual framework for developing assessments and meaningfully interpreting and using feedback to inform teaching and learning. A timely sideline to the overall discussion, it involved the 'science' of constructing and using a scale. The full text is included here because it is also relevant to some important discussions in Section 4.

When we are asked to measure properties or constructs (such as, mathematics knowledge and skill or leadership potential) in the social sciences, we are being asked to solve a measurement question. We can get some ideas of how to do this from the measurement experiences in the physical sciences.

Lots of people say that it is much easier to construct measurement scales and measure properties in the physical sciences because you can actually see height and weight. However, this is not always the case. For example, we cannot see heat; yet we can measure it. This is done by making a link between heat and the manifestation of heat (increased movement of molecules causing expansion of mercury up a capillary tube) that is calibrated in degrees Centigrade or Fahrenheit to form a temperature scale (thermometer). Once the scale has been constructed measurement is possible using the scale.

One of the advantages of measurement in the physical sciences is that it is generally objective and independent of the location in which it is used. This means, for example, that temperature can be interpreted without reference to the particular thermometer used for the measurement. It also means that different people can use it in different locations and on different occasions and the measures are comparable.

A second advantage is that measurement is always referenced directly to the scale and as a consequence the results can be more meaningfully interpreted in relation to what is measured by the scale. There is a direct relationship between the amount of the 'thing' being assessed and the scale. For example, the further the mercury is along the scale, the more heat energy there is.

These features are highly valued in measurement in the physical sciences and would be just as highly valued (although rarely present) when measuring in the social sciences.

The challenge for the social scientists (which includes teachers in this case) is to develop scales that enable measurement of the constructs that they need to measure. While the challenge is demanding the principles underpinning the development of measurement scales are the same and the advantages that accrue from using such scales for measurement in the social sciences make the efforts well worthwhile.

(Tognolini, 2005, pp. 2–3)

Reg Allen (2005) and Hilary Hollingsworth (2005) talked about data in a way that was different from most of the papers. Hollingsworth described the use of videotapes of classroom practice as a mechanism for teacher training, for general feedback about teachers, and individualised feedback to teachers. Whereas Hollingsworth described the use of records of 'teacher work', Allen argued the value of analysis of 'student work', which he used in the broadest sense rather

than just in the sense of culminating performances or assessment pieces. Student work is a window into what actually happens in schools and classrooms. Student work is enacted practice as distinct from what is said to happen – intention, wish-fulfilment or social desirability – or what is inferred from test data to have happened.

Ingvarson (2005), Meiers (2005) and Cahill (2005) each talked about the use of data to guide teachers' professional development as part of a broader reform strategy. Some of this work is treated in more detail in Section 4.

Conclusion

All these papers emphasised the use of data to support learning at practitioner or operational level. There is also, however, an argument for a meta-analytic approach. Many of the papers indicated how data had been used in particular cases. But more is needed. We need evidence about whether data *are* used more generally and that, when used, whether or how those data *do* support learning. Rigorous and sophisticated use of data may still be the realm of the few rather than the many. We need evidence of the everyday use of data by practitioners in ways that enhance the learning experience.

Dichotomies

Section 4 comments on a range of dichotomies found in current discussions about data and evidence, and suggests some alternatives for managing the discussions more effectively. By including references to additional international work, this section extends the review from Section 3, which was primarily concerned with the ACER 2005 Research Conference papers.

The term 'dichotomy' generally refers to the separation of something into two divisions that differ widely from or even contradict each other. It is also the formal term for the first-quarter and third-quarter phases of the Moon when a half-moon appears because only half of the Moon's surface is illuminated by direct sunlight. The power of this analogy is that it reminds us that the illuminated portion of an object tells but half the story. When the object of the story is assessment data, nothing less than the 'whole' will do.

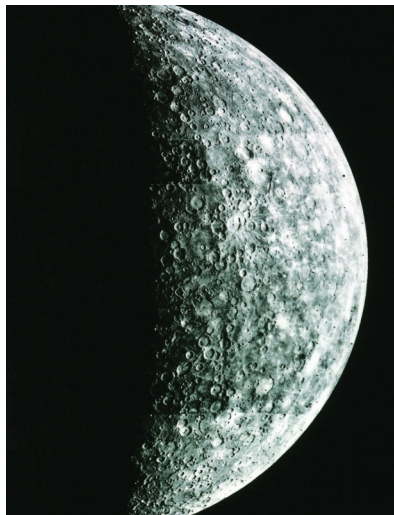


Figure 4. Photograph of the Moon in first-quarter phase

One common weapon in the handling of any change process is the false dichotomy. The weapon wielder proposes a solution to a problem which involves choosing between two seemingly opposite methods, as if they were mutually exclusive absolutes. The script usually goes something like this: An 'old' method is examined and found lacking. A 'new' method, which is asserted to be clearly superior, is proffered. The replacement of the old with the new is advocated.

When the 'happy clapping' is over, what the observer often sees is that *both* methods can be or have been traditionally used effectively. Furthermore, it is often the case that the two methods are complementary. Some obvious education examples are the teaching of reading and the replacement of the teaching of arithmetic with the teaching of problem solving.

In the heat of a discussion about these issues, it is easy to forget that learning arithmetic versus acquiring mathematical skills (i.e. being able to 'do maths') is not, and never has been, an either/or proposition in the teaching of mathematics. Ellis (2005) makes a similar point about there being no single-strategy solution for effective instruction for students (with or without learning difficulties). She advocates 'bridging the gap between constructivist and direct instruction methods' (Ellis, 2005, p. 46) as the resolution of that particular false dichotomy.

These examples are not provided as an introduction to any debates about the teaching of reading or the teaching of mathematics or teaching students with learning difficulties. They are provided simply as concrete examples of the false dichotomy by way of introducing the topic of this section.

The author of this review paper, in preparing for her writing task, was struck by the way apparent dichotomies populated the education literature. Drawing on this reading, she identified four dichotomies, which, to a greater or lesser extent, were also manifested in the conference presentations or were inherent in the research being reported upon. The four dichotomies, which are then elaborated, are listed below.

- 1 Assessment for learning versus assessment of learning
- 2 Student work versus test score
- 3 Progress of individual student versus success of school
- 4 Sociology versus psychology as an explication of success

Figure 5. The four apparent dichotomies

The lack of acknowledgement of an alternative discourse, as observed in many of the discussions, in much of the literature and at the ACER conference and other conferences, could be a function of a genuine lack of knowledge, often attributed to a lack of time for reading or to rigid specialisation in one's studies. The lack of acknowledgement could also be a function of selective citing of the literature, sometimes because one holds a certain ideological position, sometimes because one subconsciously absorbs the flavour of the last-attended conference or seminar.

The connection of Section 4 to the overall intention of this review is to identify the main discourses in current discussions about the use of data to support learning, with a view to persuading researchers and policy makers to:

- agree that the dichotomies do exist
- treat the dichotomy less as a polarity and more as a continuum
- consider what might be suitable labels on each of those continua.

Crucial to an understanding of how dichotomies work is the notion of paradigm, because it is often the case that a dichotomous situation arises in conjunction with a paradigm shift.

Paradigm

'Paradigm' has become an important technical term since the publication of *The Structure of Scientific Revolutions* (Kuhn, 1962). Kuhn's thesis was that normal science operates within a largely unquestioned framework governed by fundamental theoretical models or 'paradigms'. These ruling paradigms determine the way in which experiments are designed and observational results interpreted. Once a theory gains the status of a paradigm (e.g. Darwin's principle of natural selection by survival of the fittest), it remains unchallenged until a scientific revolution

occurs and it is overthrown in favour of a new paradigm (such as occurred in the switch from Newtonian to Einsteinian physics).

In expressing his view that the scientific method had moved on from Newton, Einstein (1916) wrote:

Newton, forgive me; you found the only way which in your age was just about possible for a man with the highest powers of thought and creativity. The concepts which you created are guiding our thinking in physics even today, although we know that they will have to be replaced by others ...

(Einstein, 1916, cited in Schlipp, 1949, p. 31)

When a paradigm shift occurs, even old established observations and experiments change their significance, and this resembles a Gestalt switch in the perception of an ambiguous figure. Figure 6 is a classic example of an ambiguous figure–ground situation. When looking at the drawing, its ambiguity forces us to shift our attention from the vase or the faces to see one thing or the other. Normally, we perceive one aspect of an event as the figure and the other as the ground. In Figure 6, there is no true figure and ground. It is a drawing that pretends to be an object. The figure–ground phenomenon is a metaphor for paradigm shift. But the fact that one pays more attention (or gives all of one’s attention) to the metaphorical vase rather than to the metaphorical faces does not preclude one from being able to shift back and forth. It may not be possible to take both stances simultaneously (the vase or the faces will be the ‘ruling paradigm’), but it is possible to reference the alternative perception.

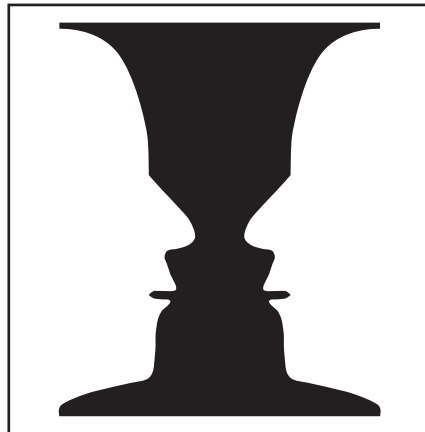


Figure 6. Illustration of the figure–ground phenomenon

We now turn from psychology to biology and then to physics to two episodes in the history of science which are pertinent to a discussion of contradictory positions. Darwin’s (1872) theory of evolution, which was challenged in the well-known 1925 Scopes Monkey Trial, has again become a hot topic in 2005 owing to the emergence of ‘intelligent design’, an anti-evolution belief that asserts that naturalistic explanations of some biological entities are not possible and such entities can only be explained by intelligent causes. It was newsworthy not so much because of its content, but because its advocates maintain that it provides empirical evidence of the existence of God or super-intelligent aliens, and so should be taught in the science classroom as an alternative to the science of evolution. Opposition to ‘intelligent design’ rejects its self-definition as scientific and therefore the public discussion has in part been about what constitutes real science.

In physics, wave–particle duality holds that light and matter can exhibit properties of both waves and of particles. The idea is rooted in a debate over the nature of light and matter dating back to the 1600s, when competing theories of light were proposed. It is now established that small objects, such as atoms, have both wave and particle nature, and that quantum mechanics (established in the first half of the 20th century) provides the overarching theory resolving this apparent paradox.

The point is *not* that contradictory positions exist or that there is only ‘one truth’. Rather, the point is that the existence of more than one position should be acknowledged (and at the very least, known about). Acknowledgement should be required, even if the alternative position is ridiculed by the other side of the debate (as in the example about Darwin) or if the seemingly contradictory positions are eventually deemed to be complementary (as in the example about Einstein).

In the text that follows, four assessment dichotomies commonly displayed in research work and in practitioner conversations are described. Then, for each of them, alternatives for more effectively managing assessment discussions are suggested. The intent is to provide support to users of data. In thinking about using data to support learning, practitioners and policy makers should ensure they keep open the possibility of more than one paradigm operating. They also need to ensure that they are alerted to, or do not fall for, the four apparent dichotomies, previously listed, and now to be analysed in detail.

Four apparent dichotomies

Dichotomy 1: Assessment for learning versus assessment of learning

There are many instances in the literature (and in conversations) of the notion that assessment *for* learning and assessment *of* learning are dichotomous. In contrast to this position, the author intends to posit that they are different rather than dichotomous.

The term ‘assessment for learning’ was coined by the Assessment Reform Group (ARG) (1999) in England. Their work highlighted the value² of assessment *for* learning as compared with assessment *of* learning. In their view, assessment *for* learning is one of the most important purposes of assessment. It is not the only purpose and is to be distinguished from assessment *of* learning, which is carried out for the purposes of grading and reporting.

There was also Earl’s (2005b) work in the United States of America on assessment *as* learning. She highlighted that attaching importance to classroom assessments is a valuable strategy for teachers in improving student learning on an everyday basis. Such attention to classroom stratagems for supporting learning has contributed to the prominence of Earl’s (2005b) work.

Since 1999, there has rarely been a conference paper in the field of assessment that did not somewhere make passing reference to assessment *of* or *for* or *as* learning. In the battle to make sense of it all, we might parody Churchill’s intonations about another battle: Never in the field of educational assessment was so much written by so many about so few prepositions³. But despite this levity, it is a serious matter, for much hangs upon the discussion, as we shall see.

Formative and summative assessment

Some clarification of the essential characteristics of two types of assessment will provide a useful context.

Formative assessment occurs when assessment, whether formal (e.g. testing) or informal (e.g. classroom questioning), is primarily intended for, and instrumental in, helping a student attain a higher level of performance. Formative assessment occurs prior⁴ to summative assessment; its purpose is partly to guide future learning for the student. Some authors represent diagnostic assessment as a component of formative assessment; some consider it a distinct form. (In practice, the purpose of diagnostic assessment is to ascertain, prior to instruction, each student’s strengths, weaknesses, knowledge and skills. Establishing these will permit the

² Black and Wiliam’s (1998) review of research into classroom assessment showed that assessment for learning is one of the most powerful ways of improving learning and raising standards. Current research is adding further evidence in support of this claim and the empirical evidence is underpinned by theory from the psychology of learning and studies of learning motivation.

³ Strictly speaking, ‘of’ and ‘for’ are prepositions, whereas ‘as’ is something else.

⁴ Although in some practices, formative judgments contribute to reported results.

teacher or instructor to remediate students and adjust the curriculum to meet each student's unique needs.) Because the primary purpose of formative assessment is feedback to the learner, it is often ungraded and, by definition, low-stakes. Formative assessment is deemed to be assessment for learning.

Vital in formative assessment is the role of teacher as mediator. According to Sadler (1998):

Formative assessment does make a difference, and it is the quality, not just the quantity, of feedback that merits our closest attention. By quality of feedback, we now realise we have to understand not just the technical structure of the feedback (such as its accuracy, comprehensiveness and appropriateness) but also its accessibility to the learner (as a communication), its catalytic and coaching value, and its ability to inspire confidence and hope.

(Sadler, 1998, p. 84)

Summative assessment occurs when assessment is designed to indicate the achievement status or level of performance attained by a student at the end of a course of study or period of time. It is geared towards reporting or certification. Summative assessment is most commonly characterised as an assessment instrument, usually conducted at the end of a term, chapter, semester, year, or the like, the purpose of which is evaluative. The Higher School Certificate (HSC) examinations in New South Wales and the Queensland Core Skills (QCS) Test are examples of summative assessments that are reported publicly though they differ in style and purpose. The HSC in NSW is an example of a subject-specific summative assessment, whereas the QCS is an example of a trans-subject assessment. Some authors represent summative assessment as an exit level of achievement in a course of study, where the final result is derived from continuous assessment. Because the primary purpose of summative assessment is reporting/certification, it is necessarily graded and, by definition, high-stakes. Summative assessment is deemed to be assessment of learning.

There have been extensions to the definitions of these two types of assessment. Gipps (2002) talked about 'assessment for reporting' and Forster (2005) introduced the expression 'reporting for learning'. It would seem that Gipps (2002) is underscoring the purpose of summative assessment while Forster (2005) is making a strong yet subtle statement about the potential of summative assessment to enhance learning – a full realisation of which will lead us directly back to the title of this review.

The following table shows the permutations and combinations of prepositions (in italics) and gerunds (in bold type) in some of the literature on the purposes of assessment. The significance of the rows labelled 'Stereotype' and 'Ideal' will unfold during discussions throughout this section.

Table 3. Comparing and contrasting assessment purposes

Source	Description	
ARG (1999)	Assessment <i>for</i> learning	Assessment <i>of</i> learning
Gipps (2002)	Assessment <i>for</i> learning	Assessment <i>for</i> reporting
Forster (2005)	Assessment <i>for</i> learning	Reporting for learning
Stereotype	Formative	Summative
Ideal	Formative and summative	Formative and summative

While it is not necessarily made explicit by Australian speakers and writers who extol the virtues of assessment *for* learning, the implications are clear. They are telling us that assessment for learning is important and good, that the education community has become obsessed with reporting on student (and school) performance, and that they believe we would be better off

focusing our energies on the feedback loop to assist students in their learning, rather than on informing others about student learning. This position is often seen as being ‘anti-testing’, even anti-reporting-publicly of assessment results. Although this new refrain about assessment for learning that has emerged in recent times might have weight in the United Kingdom, with its national curriculum and national assessments at key stages of schooling, it does not make so much sense in Australia where there is no national curriculum and where there are only some States and Territories with mandated statewide assessment in subjects or Key Learning Areas (KLAs) in Years 1 to 10. (This statement is less correct since the introduction of national testing, at Years 6 and 10, in Science, Civics and Citizenship, and ICT on a three-year rotational basis on a sample of students in each year.)

The case being put here is not that formative assessment has been overrated but that summative assessment is being blackballed. Underpinning the fashion to put assessment *for* learning in the limelight and assessment *of* learning on the ‘far side of the Moon’ is the assumption that assessment for learning serves a higher purpose. Teachers, on hearing all the comparing and contrasting of these two different assessments become so paralysed with fear about the real purposes of assessment that they sometimes lack confidence in designing their own assessment tasks or instruments, as well as in the use to which the results from these tasks or instruments may be put. Alternatively, they take the emphasis on possible negative qualities of summative assessment as an excuse to denigrate the testing movement per se and to resist any form of external assessment in the compulsory years of schooling. This was most surely not the intent of ARG (1999), Earl (2005b), and subsequent related work. That body of work reminds us of an underused, yet effectual, component of the educational process.

Some of us take the position that no distinction is necessary between assessment of and for learning in their content or conditions. Arguably, all assessment is assessment of learning. Assessment results may be used for a variety of purposes. Of course, the most productive of these purposes is the promotion of further learning, the supporting of learning.

It does not necessarily follow, however, that we should turn our backs on accountability, for that purpose may also be to support learning. Tognolini (2005) is not a strong advocate of making much of the difference between assessment for learning and assessment of learning. He is more interested in what works, what best achieves the purpose of improving teaching and learning.

In my opinion assessment is the collection of information for a purpose. I would not see a case, even in high-stakes examinations, where the assessment information that is collected should not be used to inform teaching and learning. Consequently, all tests should provide data at different levels of generality to inform the teaching and learning process. More importantly, I believe that students, teachers and parents should be taught how to interpret data themselves. As I have talked about feedback and the use of data around the world, teachers generally say to me, ‘How am I expected to use the reports to give feedback to 40, 50 or 60 students in my class?’ I am always perplexed by this question. Surely rather than go through each of the reports with the students it is more appropriate and efficient to teach the students themselves to analyse their own performance and tell you, the teacher, what they did wrong on those items they might have been expected to get right and what they might do in future to ensure that they get such items correct. In this way the learner is involved in the process. However, in order to ensure that this process works effectively the students need to be provided with a context or reference frame within which they can interpret their performance.

(Tognolini, 2005, pp. 17–18)

Others have taken up this theme. According to Stobart (2004), effective assessment encompasses the dual and related goals of formative and summative assessment. It follows, therefore, that assessment for learning and assessment of learning are not assessments that have to develop independently of each other. They both focus on improving learning, particularly those deeper forms which can give rise to active and continued interest in learning.

Wiliam (2004) argued that the terms 'formative' and 'summative' apply not to the assessments themselves but to the functions they serve and, as a result, the same assessment can be both formative and summative. But Tognolini, Stobart, and Wiliam form a small group with few published adherents. Most work in the field suggests that to practise one is to ignore the other.

It might be an oversimplification to categorise formative and summative assessment as assessment for learning and assessment of learning respectively. It might be possible to give or receive a comprehensive briefing on effective assessment without the need to employ either term. But until further research on these possibilities is undertaken, this review is forced to use the generally accepted terminology. So the author will continue to employ these terms when describing the juxtaposition of teacher-based classroom assessment and standardised statewide (or nationwide or worldwide) assessment. This discussion follows.

A superficial reading might characterise teacher-based classroom assessment (written or oral or whatever) as assessment for learning and standardised statewide assessment (testing or common assessment tasks, be they either paper- or computer-based) as assessment of learning. In a sense this characterisation derives from experience; it often results from relatively informal assessments that are used by teachers as the basis of discussion with students of their achievement or progress at class or individual level. But effective classroom assessment can also deliver comparability of meaning of reported results, if appropriate standards assurance processes are in place, as for example in externally moderated school-based assessment.

Standardised statewide assessment (for which results are invariably reported to parents and the education system, not (directly) to the students) can improve teaching and learning if both teachers and administrators take advantage of the information that the data have to offer and come to the view that 'inconsistency in performance ... becomes a puzzle to be solved' (Moss, 1994, p. 8). There is no rule that says that assessment for reporting cannot also be used in the feedback loop to students. It is only the negative, though powerful, stereotype of standardised testing (or common assessment task) which asserts as incompatible the use of results which appear on reports and productive conversations with students about their learning (see Table 1 and the discussion that precedes it).

Obviously, standardised testing and common assessment tasks can improve the comparability of results reported to parents and the system about the quality of the learning that has occurred, and under what conditions the evidence of the learning was gathered. Thus, it can also support learning. The problem seems to be that few people in Australia are willing to argue Popham's (1987) case⁵ for the 'merits of measurement-driven instruction' (translated for the Australian context as the 'positive backwash effects of assessment on curriculum and pedagogy'). This problem persists despite local and recent research (Queensland Department of Education and the Arts (DEA), 2004) which showed that programmatic emphasis on assessment in an assessment-resistant culture (Queensland Years 1 to 9) 'strongly influences teacher behaviour [and] has an impact on curriculum and pedagogy' (DEA, 2004).

Summary of, and ways of moving forward with, Dichotomy 1

Opinions may be polarised on the matter of assessment for learning and assessment of learning, but Dichotomy 1 is a false dichotomy; the polarities exist only in the minds of the beholders; they are not inherent. There are strong emotional attachments displayed in the literature and the field, which render resolution of the dichotomy difficult. By virtue of these strong feelings (particularly in the field) we need to replace this angst with a more constructive discussion about the criteria for effective assessment. The criteria for effective assessment need not

⁵ The author of this review paper is aware of the direction of Popham's later writings but still endorses the argument as it stood in 1987.

vary, whatever the purpose of that assessment. This is the reason for the row heading 'Ideal' in Table 1. And this is the way to move forward.

The continuum called for at the beginning of Section 4 to replace the dichotomy should therefore be labelled 'effective assessment', for, by adopting this new purpose and title, both the differences and similarities in assessment *of* and *for* learning are ignored or eliminated.

Because datasets are so important in our system of decision-making, Section 2 of this review exhorted us to grapple with the concepts of reliability and validity. Axworthy (2005), however, informed us that these are not attractive concepts for teachers. In order to maximise the efficacy of summative *and* formative assessment and to manage our discussion about assessment more effectively, the following factors should be considered: rigour (in assessment systems and assessment instruments); accountability; credibility; and authenticity. These four factors are no more than another way of talking about reliability and validity in all their guises. Our energy should primarily go into designing effective assessments and critiquing, at both design and item level, instruments developed and administered by others, rather than into arguing about the relative merits of different purposes of assessment. The purposes outlined as formative and summative are both legitimate. Thus, both kinds of assessment can be used to support students in their learning, if they are properly undertaken.

Unless a detailed review is commissioned, it is difficult to become aware of the persuasive nature of the dichotomy or of the way that some of the valuable messages from ARG's original work have been transmogrified. Research is needed in Australia into teachers' and policy makers' attitudes to, and understandings of, the issues raised in the analysis of what has been called Dichotomy 1 in this review.

Dichotomy 2: Student work versus test scores

Dichotomy 2 relates to the relative value and accessibility of test scores and student work as evidence of student achievement. The author intends to put the general case that student work is a neglected source of data for research purposes (not to mention for assessment and reporting purposes). It is the intention of the author to create a better understanding of some of the anomalies associated with the levels of use and acceptance by those in the field of student work (including performances) as evidence of achievement.

It is important to record at the outset that a high level of interest has been generated in performance-based assessment as manifested by more than ten years' coverage at conferences and in journals. A prominent view is that changes in assessment using more complex, meaningful and integrative performance tasks will improve education (Baker, O'Neil, & Linn, 1993). Although interest in performance-based assessment remains high (at least in the discourse), our knowledge about its quality is low. The analysis of student work could contribute to this knowledge base.

Defining terms

Before discussing the second apparent dichotomy, it is necessary to define two terms that require careful separate consideration.

The term 'student work' is taken to mean the collection of the substantive products of student learning in and/or across curriculum areas, composed in response to assessable features of an assessment task or examination question. The products of student learning may include artefacts, performances, project work, answers to examination questions and so on. Student work is the outward and visible sign, a demonstration, of learning. It is *primary* evidence of achievement of one sort or another.

A 'test' is taken to be an assessment instrument constructed by persons technically trained in achievement/aptitude test design and statistical methods. Its items have been thoroughly trialled beforehand, and the test is accompanied by norms or standards of performance that enable interpretations to be made about candidate performance. Obvious examples are the national sample assessments in Science, Civics and Citizenship, and ICT for Australian students

in Years 6 and 10 with results reported against national benchmarks. Test scores are *secondary* evidence of achievement of one sort or another.

Amid the hype surrounding the use of data to support learning, it can be easy to lose a sense of what educators and policy analysts are collecting data about. Figure 1 in Section 1, which is a framework for locating data in the student-learning environment, has test scores and grades as outputs. These grades could come from an overall assessment of work done during a course of study, or from an aggregation of marks on examination questions, or from student responses to an assessment task completed over an extended period of time. Table 1 in Section 2 includes tests of various types, performance assessments and projects as potential sources of data about student achievement. And these performance assessments presumably deliver student work as defined above.

References to test scores, their antecedents, consequences and correlates abound in the research literature but there is a dearth of information about student work in spite of a shift towards assessment tasks that, according to Shepard (1991), emulate the kind of process-based higher-order tasks thought to represent good practice in aligning curriculum, pedagogy and assessment. The dichotomy being analysed is not, however, about the formats used for assessment; it is about what students produce in response to any assessment task in any format and how that is, or should be, marked/judged by teacher-assessors or examined by researchers.

It should be noted that, just as test scores can be made available to teachers as well as to researchers and government agencies, so student work can be made available to researchers as well as to classroom practitioners. Responses such as, 'It's too hard and/or expensive to collect', 'Teachers/schools and examining bodies won't part with it' are insufficient justifications for student work being ignored as a research option more readily than any other.

At another level, student work can be an individual's test responses (as opposed to test scores), and much attention is given to that topic in the research literature. Holmes-Smith (2005) addressed this topic in his conference paper, by using student response reports from Victoria's statewide literacy and numeracy tests as a diagnostic tool and by analysing response patterns as an explanatory tool. For example, there was one particular multiple-choice question in Year 3 mathematics where 39% of students in a particular school chose the same (incorrect) distractor from a list of four options as their response. This indicated that a large proportion of students in that class from that school did not fully understand the concept being tested. The data certainly gave teachers at this school some powerful information about a concept not well understood by their students. The next valuable conversation to be had in that school's staffroom might be one about pedagogy, or it might be about the mathematics syllabus or it might be about the performance of the students in that 39% grouping on other parts of the test. The possibilities are endless in the quest for finding an answer to the question, 'Why is it that so many of our students got that question wrong in that particular way?' In the discussion that follows, student work is restricted to students' extended responses rather than to the letter-code for a closed response.

Low research profile of student work

The plausible explanations for the sparse references to student work in the research literature have been alluded to; most of them logistical and financial rather than methodological or philosophical. Excuses given are, for example, difficulties in collecting student work at the school site, difficulties in gathering student work in a central place, difficulties identifying and assembling judges, and difficulties in accessing appropriate quantitative models for the analysis (e.g. pairwise comparisons, David, 1987) and so on.

Although the ACER 2005 Research Conference adopted a broad interpretation of 'data' as outlined earlier, the data described in most presentations were measures of student achievement, usually in the form of test data. Of the 18 concurrent papers, only three referred to anything other than test scores. Of the papers that were about student achievement measures, only one of these had student work as its central platform.

Techniques using student work as direct and visible evidence of achievement, of the repertoires of practice of students and teachers, provide a powerful opportunity for teachers and schools seeking to improve the learning of the students they have.

(Allen, 2005, p. 87)

Allen's (2005) paper on student work, similar to the paper by Ingvarson (2005) on professional development (discussed in Section 5), stands out because it tells us something that challenges the status quo. To revisit the geological metaphor used in Section 2, the data in this case are like 'geological upheavals in a landscape' because they 'mark our views' about collecting and analysing data as *anything but* 'static and unchallenged'. The story they tell might not be what was expected. This is not unlike another set of data, that from the PISA study, which matches the geological metaphor. It is relatively rare that international comparative studies in education such as PISA yield information that challenges the accepted wisdom about high-achieving countries – although it must be said that the analyses underpinning the *schadenfreude* related to Germany's (surprisingly low) results and the simplistic explanations of Finland's good results are raising comment in the field (e.g. Goldstein, 2005).

This dichotomy, however, is not about the interpretation of PISA results; it is about obtaining evidence of student achievement. The crucial evidence of what students know and can do, and of how well they know it and can do it, is the (non-trivial) work that students produce in the classroom or under examination/assessment conditions. Apart from providing the primary evidence about student learning, student work also provides other sorts of useful evidence. For example, in the inspectorial sense, student work tells of the extent to which the intended curriculum has been enacted at the school or classroom level; and, in the standards monitoring sense and marker monitoring sense, it tells of the extent to which students who are awarded the highest available grade are students whose work actually exemplifies the features of the highest standard as described on paper. These other sorts of evidence are all important in supporting student learning because improvements in student learning depend on the enhanced learnings of other players described in Figure 3 in Section 3.

Allen (2005) paints a picture of 'teacher as professional', as opposed to 'teacher as technician'. He sees a professional who seeks a broad and deep understanding of what and how students learn, and who sees the student as a whole person, living in and across a time and place, and embedded in cultures. He paints a picture of student work that is accessible to classroom practitioners, assessors and researchers alike and that provides direct, visible and complete evidence of the knowledges and repertoires of practice that the student has acquired.

Allen (2005) coins the term 'point-out-able' to describe the features of student work that provide the evidence of achievement. It is the desirable features that are rewarded by markers/raters and, therefore, the assessment process sends powerful messages to teachers, students and parents about what counts. These features can also provide the basis for valuable conversations between teacher-assessors about the nature of the students' classroom experiences that might have led to such products (of learning) because, as Fullan (1993) says: 'Assessment and pedagogy feed off each other'.

The picture of student work referred to earlier not only shows the extent to which intentions in written documents have been translated into practices and experiences in the classroom; it also shows what it really means to be 'doing school' in a particular place at a particular time in a particular culture.

Finally, those who believe that test scores generate sufficient data about achievement and who advocate that test scores generate more or better data than does student work, seem not to have considered the last decade's assessment fashion, that is, authentic assessment. When working properly, authentic assessment, by definition, involves students in using relevant and useful knowledge, thinking and practical skills. (Not all assessments have or should have this intent or purpose, of course.) When the assessment is authentic, students experience the task

as it could be carried out in a non-school environment, the range of response modes is broad, and the skills developed in other subject areas are enhanced.

The observation from Section 1 that ‘the more things change, the more they stay the same’ is again relevant here. There has been tremendous change in the range of available methods for assessing student achievement but there remains an appetite for, and reliance on, pen-and-paper tests. Indeed, the combined effect of the accountability age and the technology boom have enabled test administration and marking and collation of assessment data in ways that we could only dream of five years ago. This combination of accountability and technology has only reinforced the benefits of multiple-choice testing, benefits such as high reliability, fast data capture and instant yet sophisticated analyses.

Computer-based testing has extended these benefits even further. R.E. Bennett (2000) maintains that technological advances, in particular the Internet, will revolutionise the business and substance of large-scale assessment just as it has revolutionised commerce and social interactions. This is already manifest in assessment regimes. Computer-based assessment of students can and should extend the scope of student responses that are available for scrutiny and research. Such assessment products can in fact be included in the definition of student work. There is a need for the methodical collation of, and reflection on, existing research into the use of student work (written, oral, and electronic) as data about student learning. Only when this is achieved can the benefits of technological advances and performance-based assessments be simultaneously shared by all interested parties.

Summary of, and ways of moving forward with, Dichotomy 2

Unlike the case of Dichotomy 1, with Dichotomy 2, opinions *are* polarised. Researchers are almost completely ignoring student work, with the emphasis being almost exclusively on test scores as evidence of student learning. (It is possible, of course, that they are not ignoring it but are not willing to suffer the logistical, financial and analytical complications that accompany an investigation of it.)

Also, unlike Dichotomy 1, Dichotomy 2 is a real dichotomy, because student work and test scores are different sorts of evidence – primary and secondary respectively – each having its own admirable qualities. By virtue of the prevalent assumption that there is only way one of doing things, we need to redress this imbalance with a consciousness-raising exercise. The continuum called for at the beginning of Section 4 to replace the dichotomy should be labelled ‘nature of the evidence of learning’. For, by challenging the status quo, the particular strengths of test scores and student work can be differentiated and exploited in the service of supporting learning.

Dichotomy 3: Progress of an individual student versus success of a school

The definition of ‘dichotomy’ as the separation of something into two divisions that differ widely from or even contradict each other has already been applied in this review to discussions about two apparent dichotomies. In relation to Dichotomy 1 (assessment for learning versus assessment of learning), the position put by this review was that it is, in reality, about artificial differences, aided and abetted by terminology. With Dichotomy 2 (student work versus test scores), the position was that, in reality, it is about the neglect by the education community of one division of the dichotomy for the other. Thus, these apparent dichotomies have both been described as indicative of the ‘widely differing’ kind rather than the ‘contradictory’ kind.

Yet the third dichotomy presently under discussion (progress of an individual student versus success of a school) appears to be accompanied by a certain amount of conflict. The author is of the opinion that this conflict is related to an arbitrary construction of the situation into one of competing ‘loyalties’ (i.e. simultaneously committing to datasets about the progress of an individual student and datasets about the success of a school). The claim, commonly made or implied in considerations of the issue, is that the *same data* cannot be used at both the individual

and group level, for investigating and reporting on student progress and school performance. This claim is no more than a distraction, but it frequently highjacks a reasonable consideration of the issue. The author will offer a description and hypothesis for the existence of the claim and for its being a distraction. Another issue that becomes entangled in the discussion about student progress versus success in school, is the argument about the merits of accountability programs. One view is that they result in greater student achievement gains; the other is that the 'cost of accountability systems is too high and the tools too weak to create real change in classrooms' (Seashore Louis, Febey, & Shroeder, 2005, p. 177).

Individual achievement data and aggregated achievement data are both portrayed as output data in the input–process–output model for locating data in the student-learning environment (Figure 1 in Section 1). Dichotomy 3 therefore reflects the trend noted in Section 1 towards studying outputs such as improvements in student achievement and school performance. It is more than ten years ago now that Gipps and Murphy (1994) applied the adjective 'recent' to the existence of more assessment in schools in England and foreshadowed the increased significance of this for students and the education system as a whole.

Usually, when there is a trend towards something, the expectation is that there is a trend away from something else. In the previous paragraph where the trend *towards* studying outputs is referred to, it is difficult to supply the expected 'and *away from* ...', in the sense of a concomitant move away from inputs. Excitement continues to surround issues such as teacher quality and training, students' backgrounds, teaching strategies, students' learning styles, curriculum design and delivery, and so on (all of which are inputs or processes related to student learning). Nevertheless, the obvious questions to be asked at a conference on the use of data to support learning are clearly questions about how measures of student progress and school success can have positive effects on student learning. These questions imply an understanding of not only how the measures are used but also how the measures are devised for use. It becomes vital to have a shared meaning for 'student progress' and 'school success', because misunderstandings about the nature of so-called successful schools seem to generate more bad vibes than almost any other aspect of the discussion. Later in this section it will also become necessary to define other terms pertinent to the discussion of school performance such as 'value-added' and 'status'. But the first discussion to be had is about students and measuring their progress.

Defining individual student progress

The general case is that, as students pass through school, their academic level increases. Most noticeably, reading ages rise, vocabularies widen and mathematical toolkits expand. This increase in achievement can be thought of as progress. The amount of progress over a given period of time differs naturally between individuals, progress over time for all students of a certain age being normally distributed.

The increase in academic level, as well as being subject to natural variation, is also a function of the environments in which students find themselves. School factors such as quality of teaching, availability of resources, and characteristics of the student body can have a significant effect on the progress of individual students. Since a student's environment extends beyond the school gates, there are other environmental factors to be considered, including student-background characteristics such as motivation, effort, health, and home situation.

Geoff Masters, in his concluding comments at the ACER 2005 Research Conference, stated that educators must be able to monitor growth, that reporting student progress against benchmarks is not sufficient as a mechanism for supporting learning. Progress can be reported in various ways. One method of presenting progress is with 'progress or growth maps', which provide a great deal of easily understood data to those who know how to interpret them. They provide information about the individual in context. The context can be the class, age-group, at any level of collation; regional, national or international. Individual progress can be reported as a 'growth map of achievement' with the individual's achievement being set in the broader context of state proficiency, as with the example provided in Figure 7.

LITERACY SCALE DESCRIPTION & NORMATIVE DISTRIBUTIONS

Angelico Jeffereson
Warra School of Excellence



Note: The indicators listed on this side of the scale have been derived from the tasks completed in the LLANS assessments. Only a selected sample of these indicators has been used to describe developing achievement in literacy.

Writes a variety of simple sentences; selects and controls content of own writing. Listens to a text and infers the reason for an event without picture clues. Uses full stops and capital letters to separate sentences. Identifies the purpose of parts of a text (eg, glossary, caption).

Recognises implied meaning in a short section of a simple written text. Reads with word-for-word accuracy, an unseen, illustrated reader with a narrative structure, varied sentences and a wide range of common vocabulary. Segments and blends to pronounce unfamiliar words correctly. Spells some common words with irregular patterns, eg, 'Basket'. Controls content in writing, eg, selects specific details appropriate to the piece, or includes some explanations, opinions or reasons.

Explains a story complication and resolution in a picture story book. Links images and text to construct meaning from own reading or listening. Reads with word-for-word accuracy, an unseen, factual early reader with a repetitive structure, varied content and some support from illustrations. Spells high frequency words with a range of patterns. Writes a piece that shows some overall coherence, eg, a sequence of events or a detailed list.

From own reading or listening, identifies and explains key events, and follows steps in procedures of a picture story book and early readers. Reads common words with difficult spelling patterns, eg, 'because'. Spells some high frequency words with common patterns. Manipulates sounds in words, eg, swaps 'm' in 'smell' with 'p' to make 'spell'. Joins simple sentences using conjunctions.

Offers simple explanations for a character's behavior, and locates explicit details from own reading or listening to picture story books. Reads unseen early readers with moderate accuracy (ie, omissions or substitutions do not consistently maintain meaning of text). Writes simple sentences that are mostly readable using phonetically plausible spelling for most common words. Lists ideas with little elaboration.

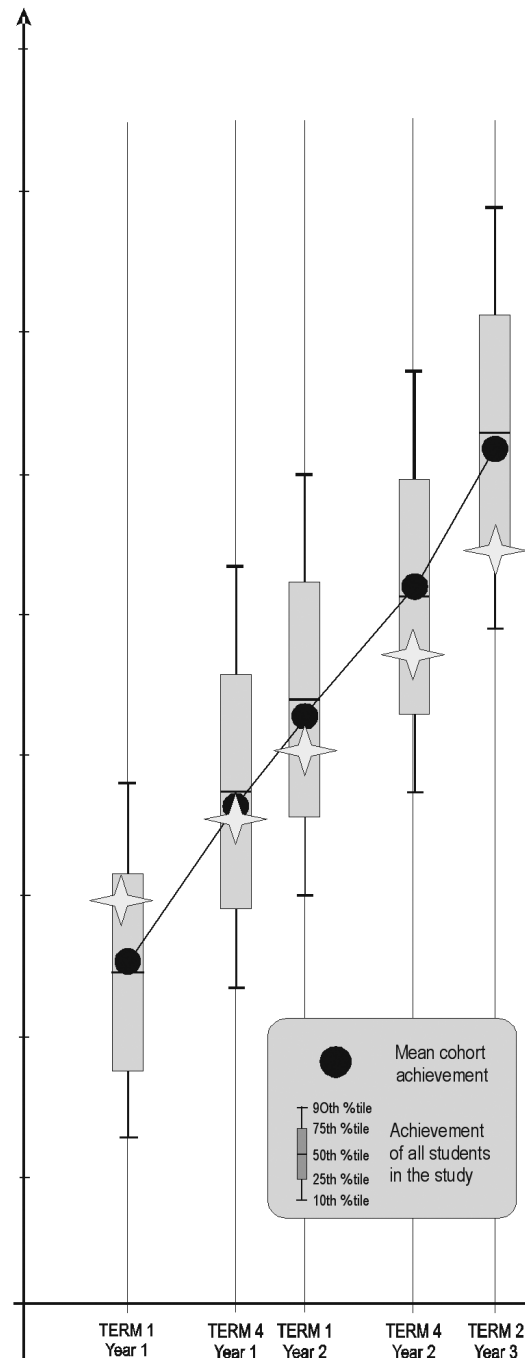
After listening to a picture story book, includes several key aspects in a retelling. Reads simple common words. Identifies all the sounds in simple words. Writes one main idea with mostly recognisable words.

Reads some very high frequency words. Recognises the same initial sounds in short words. Writes some recognisable words with spaces. Communicates some meaning in writing.

Describes an event or gives a limited retelling after listening to a picture story book. Reads a single word label by linking to the illustration. Names and sounds many letters. Writes own name correctly.

Describes the main idea in an illustration after listening to a picture story book. Identifies writing and distinguishes words and letters. Writes a string of letters or scribble.

Locates the front of a picture story book. Identifies a word.



Australian Council for Educational Research

[Source: Rowe's (2005) adaptation from Masters, Meiers, & Rowe, 2003]

Figure 7. A growth map of achievement progress in literacy showing individual group and norm-referenced criteria

The growth map in Figure 7 is explained in simple language to the report's audience via a legend included in the report. The legend derives from the underpinning box-and-whiskers plots composed by the data analysts. (The stars reference the individual's achievement.)

Rowe (2005) reported a typical comment made by parents when they received their child's progress or growth map in the format of Figure 7.

This report of my child's progress at school is great! For the first time, I have descriptions of what my child has achieved, what is currently being achieved, and what has yet to be learnt and achieved. With the teacher's guidance, I now know how best to help my child at home. Before, I had no real idea of what was expected or how to help.

(Rowe, 2005, p. 134)

Teachers commented on the utility of these progress or growth maps thus:

Using these maps, I can monitor the learning progress of each child in the class – against the norms for their age and grade levels. I can also identify what I need to do to help those children who are not progressing as well as they should.

(Rowe, 2005, p. 134)

These two quotes are evidently from people who know how to use such progress maps. In the hands of the knowledgeable, progress maps are a powerful tool.

Defining comparatives and superlatives

Clarification about terminology is again necessary. No one would disagree with the following statement: 'The *good* school, the *successful* school, and the *best* school for my child are not necessarily the same school'. The discussion proceeds on the assumption that it is not only parents of schoolchildren who would generally agree with the statement as written. But the terms 'good' and 'successful' are commonly conflated.

Analyses of the success of a school hinge on the following equation: successful school equals school that, on average, produces high-performing students, students who do well on external measures of academic achievement. The successful school in the title of Dichotomy 3 is, therefore, taken to be a high-performing school. The complications of estimating school performance permeate the whole of this discussion of Dichotomy 3.

It should be emphasised that the statement above for which agreement was canvassed includes the words 'not necessarily'. There is nothing wrong in attending or wishing to attend a successful school.

Defining the success of a school

Schools do not automatically increase the achievement level of their cohort of students over a given period of time to the same extent. That is, students at one school gain an additional advantage over students at another school. This relative advantage is known as 'value-added'. It is what the school has been able to add to the achievement of its cohort of students, given the ability of the students. Statisticians call it 'the residual' because it is that which is left over after they have taken student ability into account in their multiple regression analyses.

Thus, the residual is not just a measure of the influence of the school. As well as that, there are measurement errors in its calculation. Nevertheless, it is a respected indicator of the net effect that schools have on student progress. Put simply, value-added is the difference a school makes in the education of its students. Most league tables, however, do not rank schools this way because value-added is difficult to calculate. And this is not just a measurement issue; it is a values issue. Both issues are revisited later in this section. Much of the current thinking about estimating value-addedness stems from work in the United Kingdom (see Goldstein, 1997). There is a discussion of the work of Kingsbury and his colleagues in the United States of America in measuring the success of a school later in this section.

Changes in the USA and the UK

Recent changes in accountability regimes in the United States of America and the United Kingdom have led to similar activities in both countries for evaluating the success of schools. In the United Kingdom, the change is summarised as:

... the implementation of a national curriculum, national assessment, an external school inspection system administered by the Office for Standards in Education (OfSTED), and the publication of schools' average achievement scores on tests and public examinations. This approach is part of a general policy initiative by the British government since 1987 to promote the use of indicators by which public service institutions can be compared and their performances evaluated.

The Parents' Charter (DES, 1991), for example, requires that comparative 'league tables' of examination and national curriculum test results be published for every educational institution (schools) and Local Education Authority (LEA). The league tables consist of schools' ranking computed from average achievement scores (raw and unadjusted) on national curriculum test results at ages 7, 11 and 14 years, together with similar scores for the General Certificate of Education (16-year-olds) and A-levels (18-year-olds). The stated intention of the Parents' Charter is that these tables be used by parents to assist in choosing schools for their children to attend. However ... the British government's intention in pursuing these policies has been to meet presumed public demands of 'accountability' and the maintenance of standards.

(Rowe, 2005, p. 132)

In the United States of America, the Elementary and Secondary Act has been revised and named as the *No Child Left Behind (NCLB) Act* of 2001. The NCLB law forces states to move faster and further to improve the achievement of every student. It sets deadlines for States to expand the scope and frequency of student testing, revamp their accountability systems, and guarantee that all teachers are qualified in their subject areas. NCLB requires States to make demonstrable annual progress in raising the percentage of students proficient in reading and mathematics, and in narrowing the test-score gap between advantaged and disadvantaged students.

Reactions in Australia

Gage Kingsbury's (2005) keynote address to the ACER 2005 Research Conference, was entitled 'Benchmarks and growth and success ... oh, my!'. The paper's title indicates it is dealing with challenging, albeit North American, material, so he may have been surprised that an Australian audience reacted with the discomfort it did, given that he was no more than a messenger delivering information about NCLB. Indeed, Kingsbury (2005) pointed out inadequacies in the NCLB approach for identifying successful schools – by looking at the percentage of students who are at a certain standard (as measured by a single test) – and described an alternative approach that incorporates both status and growth – the 'Hybrid Success Model' (Kingsbury & Houser, 1997). This model did not seem to be seriously considered by the conference, yet it merits more open debate. Perhaps Kingsbury's paper tapped into the audience's negative feelings about league tables, ambivalences about definitions of successful schools, and the stated goals of the NCLB. It may also have suffered from connotations applied to the names of the inputs into the hybrid success model. The inputs, 'growth' and 'status' are now discussed.

Kingsbury used the term 'growth' in the same way as did Masters (2005) and Rowe (2005), and calculated a growth index for each student. His use of the term 'status' should not be confused with socioeconomic status (frequently used as a variable in educational research) or with status as prestige (i.e. an enviable state, associated with wealthy and successful people). Status is the name given to aggregated student performance data (e.g. averages scores on a fifth grade mathematics test). Kingsbury and Houser's (1997) model is a hybrid model because it incorporates two distinct measures, academic growth of each student and average academic performance of students in the school-group. As the academic performance of students is measured by test scores referenced to proficiency standards, a perceived over-reliance on test scores (in the USA) could be another source of the audience's discomfort during the presentation.

Kingsbury (2005) presented the following research results.

- *Schools with very similar status levels may differ greatly in the amount of growth they cause in their students.*
- *Schools will cause very similar growth for students with very different status levels.*
- *A high-performing school may not be one where you would want your child enrolled.*

(Kingsbury, 2005, p. 3)

Kingsbury reported that some schools are consistently more effective in eliciting growth for their students, regardless of the students with which they work. He is saying that some schools do a better job than others, and that this difference can be quantified. These findings, as well as having implications for recent education policy in the United States of America (see earlier reference to NCLB), are also quite confronting to schools and teaching professionals. Nevertheless, they confirm what our intuition has always told us – that schools should be judged on the basis of the cumulative value that they add to their students as a result of the students' experiences there, not on the raw ability of the students who attend the school.

Richardson (2005) tempered this enthusiasm for measuring a school's success in terms of its value-added contributions.

The value added by the school is usually estimated in terms of student and group performance above that of their peers. Yet it is rare for all academic characteristics such as ability, past performance in the subject area, teaching and learning strategies, and contextual variables such as gender and SES at student and school levels to be comprehensively measured. This level of data is just not available yet in Australia.

(Richardson, 2005, p. 124)

Rather than the more sweeping data-collection-capacity claims made by Kingsbury, Richardson made a strong case, supported by sophisticated analytical techniques, for ability-adjusted monitoring of each student's progress at regular intervals. This sort of monitoring is predicated on a measure of student ability that comes from common tasks and moderated subject assessments for each year level. Her work confirms something else that our intuition has always told us – that the learning gains of the middle- and lower-ability students might constitute the real story of a school's success.

It is a huge leap from the nuanced approach to success of Richardson (2005) and Kingsbury and Houser (1997) to the extremely narrow view of success promoted in the popular press. For this, we need look no further than at the front pages of newspapers around the country at the time of year when Senior Certificates and Tertiary Entrance Ranks (TERs) are released to students. In 2005, *The Sydney Morning Herald* produced league tables based on HSC results, as it does every year, and again cogitated about the high ranking of a certain college (Doherty, Norris, & Burke, 2005). Conclusions about the ranking of schools are reached through a misuse of data, which causes many misunderstandings, and it also causes pain. It is, of course, poor methodology to aggregate unscaled scores and expect to finish up with a fair and sensible ranking. And its power remains great, on the pages of a major newspaper.

Successful schools, like successful athletes, trumpet their victories as reported in *The Sunday Mail* of 22 January 2006.

Should schools be rated, ranked and judged on academic results? It's an endless debate but there's no doubt in the mind of authorities at [XYZ] High in Brisbane's western suburbs, where the streetside noticeboard boasts: 'Nine OP-1 scores'.

(Thomas, 2006, p. 57)

There is a certain irony here. The TER in Queensland was deliberately called the *Overall Position* (OP) to emphasise that it is not a score but a position (expressed as one of 25 bands) in a statewide ranking of OP-eligible students according to their overall academic achievement. When the OP replaced the TE Score in 1992, there had been public outrage (accompanied by disquiet from schools) about the so-called branding of students with a ‘three-digit number’ (the TE Score was expressed in the form 975, 970 etc.) and the so-called branding of schools by the high TE Scores obtained by their student cohort. This new entity the ‘OP-score’ (so called in the newspaper extract cited earlier and also by some people in the education system who should know better) is a linguistic and conceptual nonsense because, when unravelled, it would read ‘overall position score’. There is a caution that should be sounded for this school and any other school that trumpets its OP ‘victories’. The school might feel foolish the following year if it only gets a few OP-1s (or even no OP-1s). In that case, and whatever the school, that school would have to resist the temptation to attribute the poor results to student ability (or lack of it) or to student motivation (or lack of it), rather than to school effect (like making sure the boys won at rugby or providing the girls with a life-skill class called ‘Preparing for schoolies week’).

Another anecdote follows as an example of sourcing blame when students do not do so well in the TER stakes. Students who participated in the Queensland New Basics trial (2000–04) will be in Year 12 in 2006. As yet, they have not experienced Year 12 or obtained an OP. Still, at one Brisbane high school where no student obtained an OP-1 in 2005, the school attributed its ‘failure’ to the students having done New Basics. This usurpation of chronology is only necessary when influences on school success or failure cannot be acknowledged.

The recounting of these anecdotes is not to present reactions in Queensland to ‘school success’ as different from reactions in any other place; it is just that the measures are expressed in regionally specific terms. Kingsbury and other writers on value-addedness could supply us with descriptions of very similar experiences. At a general level, attribution theory (Weiner, 1985) would appear to be recommended reading. It provides a useful framework for evaluating causal ascriptions that influence motivation, achievement and performance; that is, ascriptions that influence success.

It is important to return to a consideration of the difference between good and successful and what is meant when these terms are applied to schools.

The good school

In common parlance ‘good’ means high-quality, approved of, desirable, decent, having appropriate qualities to be something or to fit a particular purpose, affording pleasure or comfort, acceptable as true or genuine. On the other hand, ‘best’ is merely the superlative in an adjectival sequence: the best school must imply the school of highest quality, the most desirable, having the maximum number of appropriate qualities to be something or to fit a particular purpose, the purpose being the education of my/your/their children. If the good school is the school that neatly fits the values of parent and student alike (sometimes one or the other unfortunately), then Kingsbury (2005) was definitely talking about *good* schools (not necessarily *successful* schools) when he said:

As long as there have been schools, there has been the question of which school is the best. From sports grounds to beautiful grounds to academic competitions, this question is discussed daily in coffee shops around the world. While it is clear that there is no ‘correct’ answer to this question, it is not for lack of trying.

(Kingsbury, 2005, p. 2)

There is an answer to the coffee-shop question about what makes a good school. The good school is the school that is a close match to a child’s educational needs (however broadly construed), generally as interpreted by a parent. Hence there might be many plausible answers to the coffee-shop question. Here a hypothetical parent’s view of the preferred school is spelt

out, for the purpose of seeing what might be included in one view. It is, of necessity, a value-laden description.

The school community at my child's school is composed of 'people like us', the principal appears to be 'well educated' and 'a good manager', the teachers are committed professionals who appear to know 'what counts' and they appear to be passionate about their vocation, the school buildings and surroundings are aesthetically pleasing, the students from that school seen on the bus/train have good manners and are sometimes seen reading a book, the school doesn't teach an English course full of the worst of today's television programs, and the school uniform and dress code bear some resemblance to a stylish and civilised society.

While this hypothetical parental view rests on one of the given dictionary meanings of 'good'; that is, as 'fit for a particular purpose', of course, it also lampoons some of the oft-unexpressed criteria for selecting a particular school. Any reader knows the variations they might need to make to this view of the 'preferred school' for some other parent and their child. All users of schools have individualised responses to the coffee-shop question. Systems may well have a different view. We are talking here of the purposes of education. Being explicit about them is important because such views and values form the basis for discussing the nature of the qualitative transformation that causes the difference between input and output data about students as they proceed through the most significant of their learning environments, the school as we know it.

Teachers as researchers

Good schools are not just places fit for students to learn in; good schools are places that are fit for teachers to teach in because teachers' attitudes influence the teacher–curriculum interaction and the student–teacher interaction, both of which are vital in the student-learning process (see Figure 1 in Section 1). A teacher might express the following sentiments, amongst many, many others, in one way or another.

- As a teacher, I would like the students in my class to do well at school (i.e. progress and achieve at the highest level).
- There are some schools where I may not want to be employed because of the way curriculum is delivered there.
- A good school is a school where I feel a sense of control.

Research is a way of giving control to teachers. The originator of the concept of teacher-as-researcher was Stenhouse (1975, 1981) who argued that curriculum research and development should belong to the teacher, and that proposals from policy makers were not diktats but ideas that the teacher should test out in his or her classroom. His vision for significantly improving education was the creation of a research tradition that is accessible to teachers and that feeds teaching.

It is not enough that teachers' work should be studied, they need to study it themselves.

(Stenhouse, 1975, p. 144)

Perhaps too much research is published to the world, too little to the village.

(Stenhouse, 1981, p. 17)

Hargreaves (1996) strongly supported Stenhouse's views in stating that:

... research gives practitioners new insights into the taken-for-granted aspects of their everyday world; it allows them to take a different angle on the familiar and to approach it for a while, with the eyes of a stranger.

(Hargreaves, 1996, lecture to Teacher Training Agency)

Teachers as policy implementers

Good schools are places where teachers are clear about what is going on. Seashore Louis, Febey and Schroeder (2005) investigated how teachers make sense of accountability policies in high schools:

School staffs exhibited the artefacts of collective sense-making, in that there was some degree of consistency of interpretation of the implications of state policy for their school.

The way in which teachers collectively interpreted and made sense of their state policies was directly tied to their willingness and propensity to change.

The collective sense-making process involved developing an understanding or interpretation of the meaning of professional control and responsibility in light of increased regulation of the curriculum.

Schools differed in the degree to which they appeared to have created structured opportunities to learn about and interpret state policies, but there was evidence that collegial conversations were significant for many or most teachers in all schools.

(Seashore Louis, Febey, & Schroeder, 2005, p. 177)

It is an interesting exercise to compare the findings of Seashore Louis et al. (2005) in the United States of America with those of Friend (2002) in Australia. Friend studied the construction and reconstruction of messages within the context of a reform process. The level of congruence between the intended messages from a central department and the school-constructed knowledge varied significantly across schools.

Some messages are not heard, others are transmogrified, but some are heeded.

These variations are influenced by the local conditions operating at each school site and the cultures that pervade the school community.

Schools with a culture that supports the school community in participating in the process of reform (which might involve them in activities like discarding traditional practices, risk taking, contesting hegemonic paradigms, and perhaps developing alternative leadership styles), where a cohesive strategic leadership supports an active professional learning community ... construct knowledge that is more congruent with the message intended ...

(Friend, 2002, p. 26)

Students and reasons for progress

Given the attribution of failure that is made against schools, it is timely to consider the status of students in this discourse. The tirade against educators is not helped by constant repetition of the mantra that teachers make the difference. A different viewpoint could be taken.

Not all, therefore, hangs on the activities of the teachers ... If the schools are failing, indeed, then we have to explore the equal possibility that it is not educators, as a whole who are necessarily at fault; rather, might the blame be laid squarely at the feet of our young?

Yet delicacy, rather than logic, might suggest this to be an indecent proposal. To entertain blaming the young for our educational situation may sound a bit like entertaining a proposal to torture the innocent. Still, it seems to be the only way to confront the educational reform movement with the logic of its own position. For if educators are fair game because of their causally central role in the learning process, then students, who are equally causally central, can hardly be spared similar attention. Fairness simply demands it.

(Ericson & Ellet, 2002)

In considering the student as a serious variable in their own learning, one should not view them as simply themselves. They are an amalgam of everything that has already influenced them, and they continue to be shaped by current influences, both internal and external to the school. Some aspects of this slant on the current dichotomy are taken up in the discussion of Dichotomy 4.

More about values and valuing

Presumably a successful school would be one that greatly enhances its students. Consider these two examples. School J may produce some 'brilliant' TERs at Year 12 but, having drafted brilliant students in Year 8, the school may not have added very much. School K may produce a good proportion of mid-range TERs from an intake of low-ability students or students from relatively disadvantaged backgrounds, and therefore these results indicate the school is enhancing those students' achievement. Exactly how much enhancement, exactly how much value is being added depends on what is defined as being of value in the first place.

As D.C. Bennett (2001) observed, value has many dimensions. No school tries to develop only a single capability in students. Schools make public statements about the array of capabilities that they aim to develop in their students. Teachers routinely attempt to develop a range of understandings and skills in their students. Measurements of value added must therefore attend to a number of different dimensions of value. So, which measures reflect which parents' values and which schools' intentions? Schools are different in that they do not all seek to add the same kind of value to students' development. So we need to measure value added in terms of a school's publicly stated aspirations for its students.

Some consequences of a school's education might take years to express themselves. We might need to measure some of the effects that unfold later in a person's life by looking at 'old boys' and 'old girls' rather than at the senior class. Although the nature of the desirable effects would depend on the values espoused by the particular school, it is very likely that social and ethical development would feature in a discussion of the consequences. Towards that end, Forster (2001) has been able to develop a set of attitudes and values scales (e.g. Conscience) that can be administered to students in the years of schooling, and to graduates five and ten years after leaving the school.

There are other measures of educational effectiveness, some of which find expression through research into teacher effectiveness. Sanders, Saxton and Horn (1997) aggregate annual gains in student achievement for students taught by a given teacher and compare these aggregate scores. In terms of Figure 1, the input data are teacher training and experience – not to mention a love of learning and the ability to pass that on, which are so hard to measure, but which parents, students and principals constantly affirm are the keys to student learning. The output data are teacher-specific student gains in test scores.

None of the above observations or opinions is intended to denigrate the notion of identifying successful schools. The issues are complex. The process done properly would be expensive. But society has to be serious in confronting the issue of whether the current arrangements are the best way to achieve more and better learning for more students. In that sense, more so than in any sense of competition between schools, measures of value-addedness or growth are examples of the use of data to support learning.

Summary of, and ways of moving forward with, Dichotomy 3

Opinions are not polarised on the matter of collecting and using data about the progress of an individual student and collecting and using data about the success of a school. Each division appreciates (or at least understands) what drives the position of the other division in this dichotomy discussion. Poised at one end of the beam of a metaphorical beam balance are those who tip the balance to datasets about the progress of an individual student. Poised at other end of the beam are those who tip the balance to datasets about the success of a school. When the beam is balanced, discussion is of a kind where there is weighing up of both sides of the dichotomy, which leads to the identification of an alternative for more effectively managing decisions about the use of such datasets to support learning. At the beginning of Section 4, a call was made to replace each of the dichotomies, especially those accompanied by conflict, with a continuum. The continuum in this case should be labelled 'good compromise in the use of data to serve the student and the State'. Garth Boomer's (1988) understanding of the importance of compromise in education is revealed in his telling of the anecdote below.

In 1978, when I was quite a young radical, I said to Jimmy Britton, as though I might never see him again, 'Jimmy, what do we need to do for the next ten years in education? What's your message to me, oh great guru and mentor?' He said, 'We need to make better and better compromises.'

(Boomer, 1988, p. 5)

It might be easier to collect data for one division in this dichotomy than for the other. But the metaphorical beam is not balanced around ease; it is balanced around acknowledgement of the fact that collecting data for both divisions is equally important. And it is crucial that educational researchers and policy makers are honest about what is being measured because this will contribute to a balanced approach to understanding the merits of collecting data on the progress of an individual student *and* data on the performance of a school.

Dichotomy 4: Sociology versus psychology as an explication of success

The title of Dichotomy 4, the final in the series of apparent dichotomies being analysed in this review, has two significant components. The first component 'Sociology versus psychology' refers to the method of the research (i.e. is it located in the sociological paradigm or in the psychological paradigm, or does it draw on both?). The second component 'explication of success' refers to the subject of the research, and is restricted in this discussion to academic success (i.e. what are the factors that explain academic success?). Research into academic success at the level of student or school generally operates within the paradigms of sociology or psychology and this thus generally determines the way in which research questions are framed, a research methodology adopted, and results interpreted. In its extreme manifestations, Dichotomy 4 can be seen as a clash of different traditions coming from two different academic camps, which live in two different worlds of factual knowledge and taken-for-granted assumptions (Segerstrale, 2000).

The operationalisation of the dichotomy, sociology versus psychology, as an explication of success, in research and in the real world is of a one-sided view; that is, the division exists but one side is either unaware of, or decides to deny the existence of, the other. Or, as is possible

according to the definition of ‘dichotomy’ that has previously been applied in this review, the two divisions in the dichotomy actually contradict each other. Contradictory status was not assigned to the divisions in any of the other three dichotomies already analysed in this review. In this section, the author explores the nature of Dichotomy 4 with the aim of convincing readers of the following.

- Researchers should continue to satisfy our curiosity, to test entirely new hypotheses, and to pour out solid data for a new world without the clashes (even controversies) alluded to above.
- The education community should not be selectively blind to the findings of such research on the basis of the paradigms in which the data were generated.
- The education community and the media, which are two subsets of the broader society, should attempt to read original text that relates to new and challenging ideas about learning, rather than accepting what the critics say/write (however plausible that might be).
- Both paradigms have merit.

Terminology

For the current analysis in this review, it is necessary to introduce into the discussion the key distinctions in the meaning of terms such as ability, aptitude and achievement.

Achievement is taken to mean the accomplishment of a particular body of knowledge and/or set of skills, usually after training or instruction. It is not the same thing as ability (capacity) or aptitude (potential). Ability (capacity) is taken to mean the current performance of a person in some defined domain of cognitive, mental or physical functioning. It relates to what that person can actually do, not what he or she might be able to do in future. Aptitude is what a person might be able to do in the future as in ‘shows an aptitude for ...’.

When attempting to do research that requires a measure of student ability somewhere in the research design, some researchers slide between these meanings and use a substitute measure (proxy in the research parlance) for ability that is no more than a reconstitution of achievement data. Since achievement data are, in turn, often used as the measure of academic success, we could find ourselves in the situation that achievement is used to predict achievement.

Research context

The *how* of using data to support learning involves collecting reliable data, making valid interpretations, reflecting on those interpretations, and acting upon that information if and when necessary. Figure 1 in Section 1 provided a framework for locating data in the student-learning environment, as input data, process data or output data. Table 1 in Section 2 provided a nine-category listing of potential data sources and linked each of those to the data locales in Figure 1. Neither Figure 1 nor Table 1 was designed to provide a framework for interpreting patterns and relationships in data.

Many researchers study the relationship between student background characteristics and student achievement. Figure 1 in Section 1 locates student background characteristics in data input, the ‘presage’ component, and student grades in data output, the ‘product’ component. Researchers might look for a link between personality variables (e.g. test anxiety) and academic success, or for a link between demographic variables (e.g. regional identity of the school attended) and academic success. Some researchers work within a causal model and investigate the correlation between ability and achievement. Others look at socioeconomic status (SES) as the independent variable in the causal relationship. Generally, there are a number of optional causal relationships to be researched, but all researchers in this situation are seeking causal relationships that will explain the data. Regardless of what is their preferred working paradigm, the majority of researchers prefer to locate the causal agent(s) in the student-learning process, which is one component of the organisational framework (see Figure 1 in Section 1). Often,

as mentioned in other sections of this review paper, they identify quality of teaching as the particular 'active' agent.

Specialist journals aside, the frequency of SES explanations in education circles far outstrips the frequency of ability explanations of achievement and between-group differences. Thus, the explanations in the educational writings that reach the public in the present era are overwhelmingly of this ilk. Possibly the frequency of ability or behavioural explanations may increase in the future, given that ability was, only recently, demystified when questions about DNA coding were answered through the Human Genome Project, which was completed ahead of schedule in 2003.

The exciting dangers of using biological data to support learning

Addressing this fourth dichotomy takes us onto dangerous ground because, in the words of Edward Wilson, 'academic theorists have paid little attention to biology ... To varying degrees they have been more influenced by postmodernism [than biology]' (Wilson, 1998, p. 214). Without wanting to inflame the debate, the author's extrapolation from Wilson's (1998) writings is that many researchers and practitioners in the field of education would argue that there is no scientifically constructible map of human abilities from which deep explanations of academic achievement can be drawn.

All educators seek answers to that elusive question about observed differences in achievement between sub-groups of the population. For example, Willingham and Cole (1997) investigated gender differences on test format (multiple-choice and free-response); Stage (1994) has noted gender differences in spatial ability (and its consequences for test design). It causes great tensions to raise questions about possible biological roots to intellectual differences particularly between sub-groups of the population (e.g. gender or race). And as a result of past misuse of data about individual differences, a UNESCO agreement in 1952 effectively banned biological research into human behaviour. But that pre-dated Watson and Crick's cracking of the DNA code in 1959 and the development of brain-imaging techniques whereby neuroscientists can illustrate that important neurochemical parameters correlate with cognitive problem solving (Nyborg, 2003). This type of research yields incredibly exciting information about the molecular basis of human nature and intelligence. Unfortunately, the overtones of eugenics have tipped over into the magic of the double helix and militate against the potential of genetics as an explanatory tool. By contrast, it is fascinating to view the public acceptance of the use of DNA in solving murder crimes. From this it can be assumed that there is also public acceptance of the concept of DNA as the carrier of all life's hereditary information.

Neurobiology is very convincing in its explanatory power, so do we in the education community really believe that not seeking answers to questions from possible genetic explanations of intellectual differences is good enough? Such information regarding the differences between individual students of the same race and same gender can guide approaches to supporting the learning of disadvantaged and underachieving students regardless of other explanations given for their disadvantage or underachievement. This would be the ultimate use of data to support the learning of the very students we so much want to protect and have flourish.

Paradigms of sociology and psychology

Sociology is defined as the study of human groups. In the broadest sense, sociology is concerned with understanding patterns of human relationships, their causes and their effects. Unlike psychology, sociology does not attempt to explain the behaviour of a particular individual under certain circumstances. Rather, sociology focuses on social trends or other influences that affect whole groups or categories of people. The emphasis that sociology places on human groups, rather than individuals, stems directly from the work of Emile Durkheim (1858–1917). He reasoned that the characteristics of a social group viewed as a whole cannot be determined simply by examining the characteristics of its individual members, nor can individuals be understood strictly in terms of the individuals themselves but as members of a particular group.

There are five main approaches within the field of psychology – behaviourist, psychodynamic, biological, cognitive and humanist. Arguably, the most contested is the biological approach in which, as already mentioned, behaviour is explained in terms of the nervous system and genetic factors. This gives rise to a major debate – the well-documented and well-known debate about nature versus nurture, which is concerned with the extent to which human behaviour is determined by heredity (nature), and to what extent it is the product of learned experiences (nurture).

Researchers of behaviour genetics collect voluminous data on families, twins and adoptions, and they study the molecular basis of intelligence. Recently, Wainwright, Wright, Luciano, Geffen and Martin (2005) in an Australian study using identical and fraternal twins, carried out a multivariate genetic analysis of academic skills displayed in performance on a standardised test of cross-curriculum skills. They found that a genetic general factor accounted for virtually all genetic variance in the component academic skills scores, and 56% and 42% of the variance in verbal and performance capacity respectively, suggesting that the factor explaining all this variance is genetic (called *g*). Another single common factor explained common environmental effects.

The modern conception of the controversial construct, *g*, which has been in psychology textbooks for a long time, can be found in Carroll (1993). More often discussed in education forums, however, is the work of psychologist, Howard Gardner (1999), who proposes that there are multiple intelligences: linguistic, logical-mathematical, bodily-kinaesthetic, spatial, musical, interpersonal, and intrapersonal. The field is large, and educational professionals need to be cognisant of the main structures and ‘messages’ as new work emerges.

Summary of, and ways of moving forward with, Dichotomy 4

In the times before the physical sciences were able to provide explanations for natural disasters like tsunamis, human beings blamed such disasters on the actions of personified gods. Now tsunamis can be explained in terms of wave motion and earthquakes under the ocean. In present times, the social sciences tend to find explanations for human behaviour in terms of society, the establishment or other personified entities. Educational sociology and psychology both aim to provide explanations for the demonstrable differences in human ability and achievement. For some educators, it might be easier to cope with the inequities in our society (such as the phenomenon whereby the glittering prizes often follow academic success), if the explanatory frame included societal and environmental factors; for others, it would be easier to cope with the inequities if the explanatory frame included biological and psychological factors.

Although it might be possible to justify the existence of Dichotomy 4 in terms of preferred mechanisms for coping with data about differences in student learning outcomes, it is not acceptable to permit one division of the dichotomy to ignore what the other has to offer in its methodology and explanatory power.

The introduction to Section 4 foreshadowed that, for each of the four apparent dichotomies under discussion, an alternative would be suggested for more effectively managing assessment discussions. This dichotomy is at a different level of generality from the others, insofar as the other dichotomies were specific constructions over the recent past by the education community in response to the issues of the era, whereas Dichotomy 4, as it is practised in the education field is part of a universal response to ways of researching and explaining human behaviour.

The explication of success in the educational process is extremely important. Sociology and psychology have their own unique methodology and epistemology, from which derive their differing approaches to the explication of success. None of the big problems in the world can be solved within one discipline. Educators and policy makers owe it to themselves and to those who are expert in the psychological and sociological paradigms to go with a multi-disciplinary approach to the use of data to support learning.

Conclusion

Section 4 identified four apparent dichotomies in our current discussions about using data to support learning. For each of the dichotomies, its two divisions were characterised and reference made to the ACER 2005 Research Conference papers, the wider educational research literature, and practical experiences in Australia. Where applicable, the model for locating data in the student-learning environment (Figure 1, Section 1) was also referenced. The analysis of each dichotomy concluded with a suggested alternative for more effectively managing assessment discussions – a continuum thus replacing a polarity.

Any discourse that dichotomises assessment for learning and assessment of learning is not helpful and should be replaced with a discourse that focuses on criteria for effective assessment, whatever its purpose. Assessment for learning versus assessment of learning is a false dichotomy, which encourages the honouring of assessment for one purpose (feedback to the learner) at the expense of assessment for another purpose (reporting or certifying) when in fact any effective assessment can serve both purposes.

Student work versus test scores is a real dichotomy, not so much because the educational community has ever put a strong case supporting one or other as the best source of evidence of student achievement, but because researchers have almost completely neglected the primary evidence, student work (written, oral, electronic) in their analyses of achievement. The silence that can be interpreted as an assumption that test scores produce the best source of evidence of student learning should be replaced with discussions about the nature of evidence about student learning.

The apparently conflicting positions taken by those whose loyalty is to data about the progress of an individual student and those whose loyalty is to data about the success of a school is an understandably emotive issue in the current era of school and system accountability. This situation is, however, presently being defused, owing to an increasing commitment of both divisions of the dichotomy to a more balanced approach to understanding the merits of the other division.

Unlike the three apparent dichotomies previously summarised, sociology versus psychology as an explication of success is not specific to this era or to education. That both paradigms are fully exploited in education in the present era is absolutely vital to our full understanding of the factors that affect student learning.

The general lesson to be taken away from the analysis in Section 4 is that there are many ways of using data to support learning and that each has its own special wonder. In acknowledging the existence of these four particular dichotomies, it is ultimately not so important whether they are real or imagined. What is important is the resulting identification of *eight* discrete approaches that produce a wealth of information that can be used to support learning, especially student learning.

Each of the eight approaches is typified (but in no way comprehensively exemplified) in the following list of pairings that were created during the analysis of the four dichotomies, but can now stand as markers of key sources for, and methodologies in, educational assessment practice and research:

- Data from classroom assessments (including observations)
- Data from continuous assessments and external examinations
- Evidence of student achievement in the form of test scores
- Evidence of student achievement in the form of student work
- Datasets about the performance of schools
- Datasets about the performance of individual students
- Sociological explanations of patterns, trends and relationships in data
- Psychological explanations of patterns, trends and relationships in data.

The implications for practice and policy of this list alone (not to mention dichotomies that other authors and readers might suggest) require a consideration of the attendant methodological, strategic and ethical issues, and this is highlighted again in Section 5.

Implications

Section 5 looks at implications for practice, policy making, and program implementation and reform in Australia of what has gone before in this review. Qualifications aside about the brevity enforced by this review and its necessarily selective coverage, the substance of Sections 1 to 4 offers insights and perspectives that can guide current practice and policy formulation.

The implications are discussed from three perspectives – the methodological, the strategic and the ethical. Examples are included for each perspective in order to temper with a reality check the readers' enthusiasm for the words and meaning of the conference theme 'using data to support learning'. Some new research and some other new material are introduced into the discussion. Section 5 concludes with a summation of propositions that the author has supported in the development of this review and with a list of advocated actions. All of these issues are attended to or have been attended to, alone or in combination, in research in some place at some time at varying levels of excellence. What is advocated here applies to the academe–bureaucracy interface and the school–research interface in Australia circa 2006.

Methodological perspective

The methodological perspective encompasses the theoretical, operational and practical components of data use. It is generally concerned with achieving optimal quality in procedural and technical aspects of data gathering and analysis. In the assessment field, examples include the use of new technologies in order to facilitate assessment, differential item functioning, multidimensionality, validity and reliability, and new concepts in designing assessment models.

At the ACER 2005 Research Conference on using data to support learning, presenters shared with participants their state-of-the-art thinking and experiences in displaying and communicating results and in monitoring trends in performance over time and years of schooling. The methodological aspect that seemed to most capture people's imagination was the use of new technologies such as computer-based item banks. Another relatively new technology, item response modelling, appears to be a taken-for-granted part of the testing movement in Australia.

Two phenomena, the centrality of the teacher and the ever-increasing emphasis on accountability, have been referenced within this review paper on more than one occasion. The related examples chosen for further discussion of the methodological implications of this review

on the use of data to support learning are the enhancement of teachers' skills in the use of data and the collection and analysis of data on the attitudes of teachers towards assessment.

Professional development

In his opening presentation at the conference, Geoff Masters stated that 'the interpretation of data requires expertise (a whole new set of skills)'. The level of teachers' current knowledge vis-à-vis an up-to-date set of skills about the use of data to support learning has been alluded to on various occasions during this review. The research and the conference proceedings indicate a need for professional development programs in techniques for interrogating student data for educational professionals.

There are many examples around the country of forays into this area; for example, a good starting point for novices exists in the approach taken in a seven-page document on interpreting attitudes to school survey reports (Victoria Department of Education & Training, 2005).

Meiers (2005), having researched the impact of the professional development associated with Western Australia's 'Getting It Right Literacy and Numeracy Strategy', was able to report on many aspects of that strategy. Of particular relevance here is her conclusion that teachers had become more reflective in their use of performance data to improve planning at the whole-school level (to a moderate or major extent): 92% of teachers in 2004 compared with 72% of teachers in 2003.

Ingvarson (2005) re-asserted the 'central importance of the content that is learned in professional development' and goes on to say that 'change in practice is more likely to be pervasive when it is informed by theory in which the educator involved has confidence' (p. 69). He related his research finding to the eighth of Hawley and Valli's (1999) nine research-based principles for the design of effective professional development:

Professional development should provide opportunities to gain an understanding of the theory underlying the knowledge and skills being learned.

Because beliefs filter knowledge and guide behaviour, professional development must address teachers' beliefs, experiences, and habits. Furthermore, specific knowledge and skills that work in one setting, sometimes do not work in others.

When teachers have a good understanding of the theory behind particular practices and programs, they can adapt the strategy they learned about to the circumstances in which the teacher is trying to use it.

(Hawley & Valli, 1999, cited in Ingvarson, 2005, p. 65)

Teachers' attitudes and expertise

In a recent survey carried out by Scott, Heyworth and Fairweather (2000), Australasian physicians self-identified the following impediments to their making better use of research data: insufficient time (74%), limited search skills (41%), and limited access to evidence (43%). Just like their medical counterparts, Australasian teachers would probably say that they do not have the time to read the research articles. Still, they always want to know the answer to the 'how do I do it?' question. Their professional self-concept seems commonly to depend on a desire to 'do the right thing by kids' (an oft-repeated expression) – and to be doing this immediately, the day after they have been 'professionally developed' (immediacy being their constant driver). For this commitment to the student–teacher and teacher–curriculum interactions (located in the student-learning process in Figure 1 in Section 1) they are to be commended. But it is possible that, in their desire to expand their repertoire of skills of the 'how to do it in the classroom' variety, they are missing opportunities to acquire knowledge of another kind such as knowledge of the theoretical underpinnings of assessment. For, when they are confident

in that area, they are in a stronger position to evaluate, make sense of, or reject if necessary, reforms dictated from above or fashions promoted by the colleagues who work beside them.

There is a dearth of research into Australian teachers' attitudes towards assessment in the compulsory years of schooling. However, the results of a recent survey (Brown, 2002) administered in New Zealand and Queensland to a sample of teachers of Years 1–9 provide insights into teachers' conceptions of assessment (DEA, 2004). Data were collected about the extent to which teachers agreed or disagreed with statements about the function and nature of assessment. Responses loaded on to four scales:

- Improving teaching and learning
- Irrelevance of assessment
- Accountability of teachers and schools
- Accountability of students.

Research reported elsewhere (DEA, 2004) includes details of significant differences between jurisdictions and of observed patterns (a desirable pattern of teacher responses about their attitudes to assessment would have relatively high scores on the Improvement and Student Accountability scales, and relatively low scores on the Irrelevance scale). One of the DEA findings is of particular relevance to the theme of using data to support learning. It is the direction of the large and significant difference between jurisdictions on the School Accountability scale (higher for New Zealand teachers) and on the Student Accountability scale (higher for Queensland teachers). The importance of this finding is that different conceptions of accountability might reflect different assessment and examination regimes in operation. This finding invites further research.

Summary comments on methodological perspective

In the current environment of increased use of student performance data for accountability purposes, it is timely to be concerned about the conduct, analysis and interpretation of data, both in large datasets like international surveys and in national and state assessments (as in the study described above). It is also timely to gain a better understanding of how teachers view assessment. The stakes are higher than ever, and the requisite demands on reliability and validity, which are at the core of the methodological perspective, are extremely high.

Strategic perspective

The strategic perspective encompasses the political and organisational components of data use. It is generally concerned with the elaboration of educational policies, the process of decision-making, and the evaluation of educational programs. In the assessment field, examples include the role of assessment in planning, monitoring and decision making, and responses to research on assessment-led reform (or any educational reform for that matter).

In advocating that the following intention should be paramount when consideration is being given to composing datasets for the users of data, Rowe (2005) included systems in the list of stakeholders. (See Figure 3 in Section 3 for the number of occasions in which the system is learning something from the data).

The utility of systems, schools, teachers, parents and students being data-informed via performance indicator feedback ... evidence for the kinds of feedback data that support learning at all stakeholder levels.

(Rowe, 2005, p. 133)

Given that, at the ACER 2005 Research Conference, one could not help but be saturated by the complex and political use of data, the first example chosen for further discussion focuses

on the responses of policy makers to research data about the players in the education delivery system, including the system itself (see Figure 3 in Section 3). The second example, discussed later but in not as much detail as the first, is the existence of evidence–practice gaps.

Policy makers and data

Ingvarson (2005) expressed concern about two aspects of educational reform. One was about the expectations of policy makers.

Policy makers can ... have quite naïve expectations about how easy it is to bring about educational change, not understanding that the kinds of change that really matter in education are not structural changes but those that build teacher capacity and professional capacity among teachers and principals.

(Ingvarson, 2005, p. 63)

His other concern was about the reality of bringing about change.

A common refrain in evaluation reports of educational reform is the lack of fit between ambitious goals for school improvement and the resources necessary to bring about significant change in practice.

(Ingvarson, 2005, p. 63)

Much of the dissonance in our systems seems to be a function of the tension between the policy makers (who often have a cynical attitude to practitioners), the implementers (those who make things happen), and the impossible political timeframes in which we operate. As Luke (2000) observed:

The internal clock of ... change has been attuned to political cycles and constrained by economies of time and scale.

(Luke, 2000, p. 29)

The literature on time-lags for changes from an educational reform process is relevant here. According to Fullan (1999), it takes about three years to flow through to practice in an elementary school and (depending on size) about six years in a secondary school.

Bowe and Ball with Gold (1992) stated that there are two flaws in conceptualising change and it is an interesting exercise to apply them to educational reform. One is that reformers often assume (or act as if they believe) that one facet of change can be addressed in isolation. The other is that reformers often assume that life begins with the moment of innovation (and then wonder why there is resistance to that innovation). Experience tells us that innovations/agents of change are related to, and build on, a history of old ones. Also change is set within, and accommodated to, the micro-political history of the institution. According to Dahl (1970), failures in reform can often be traced to the forces arrayed against the reformers, not only because of the forces arrayed against them but also because the pictures in their minds about power and influence were simplistic and inaccurate.

Freebody (2005a) queried the practical commitment of some decision makers to a research base. It is the prerogative of high-level decision makers to accept or reject research findings. This important issue is raised here so that there is no false euphoria about the status of research-based decision making.

It is important to note that a commitment to research-based decision making brings with it the responsibility to spell out any other compelling overriding or compensatory bases for decisions; that is: When will competently conducted and produced, locally relevant research not clinch the case? What constitute other potentially overriding or compensatory factors, and when will these become relevant and applicable?

(Freebody, 2005a, p. 18)

In answering his own question, Freebody leaves researchers and policy makers, bureaucrats and teachers with food for thought, in pointing out that there are other sorts of admissible evidence not just research evidence that come into play when decisions are made about how the data will be used.

It may be, for instance, that certain groups made especially vulnerable by a decision have not figured in the research data to an adequate extent, or that some professional or community groups have such an aversion, moral or otherwise, to the warranted recommendations of a research project that such recommendations cannot be adequately effected at that time.

(Freebody, 2005a, p. 18)

With even greater perspicacity, Freebody (2005b) echoes Shulman's (1988) despair about 'vague principles and maxims' in stating that, for most theory and research, we have 'unbearably generic' recommendations for teachers and educators. It is particularly crucial from the equity perspective that recommendations do not ignore the diversity of students or the significance of their particularities. It is also crucial that recommendations made in the spirit of improving student learning are stated unambiguously even if their import is not universally acceptable.

Evidence–practice gaps

In analysing the research evidence underpinning current practice, the author was struck by a number of evidence–practice gaps in the use of data to support learning – or at least in the large number of related issues that are just not addressed. Some of these were mentioned in Section 4.

Kimmelman and Kroeze (2002) have made similar observations about evidence–practice gaps in the United States of America. They wrote:

[T]his increased emphasis [on standards-based curriculum reform and assessment based on those reforms] has arguably led to improvements in school curricula and student achievement. It is somewhat ironic that they have also created a 'research implementation gap'. This gap is between those who analyse educational data and the daily practitioners, teachers and administrators who use it.

(Kimmelman & Kroeze, 2002, p. 51)

Summary comments on strategic perspective

With the current excitement about the use of data to support student learning (and the learning of all other players/agents in education), it is timely to give serious attention to situations where there appears to be a disjuncture between what the research says and what happens in practice. Such situations have consequences for the way that teachers and students spend their time. A discussion about consequences for teaching and learning is an equity issue of sorts, and equity is discussed next.

Ethical perspective

The ethical perspective encompasses the social and political components of data use (these components also thread the methodological and strategic perspectives). The ethical perspective generally calls for the exercise of integrity and rigour. In the assessment field, examples include fairness in testing and assessment, anticipating consequences of the use of data, management of risks associated with data collection and analysis, adequacy of the data upon which to base generalisations, appropriateness of data use, and undue political intervention. From the preceding list (which is by no means exhaustive), it can be seen why the ethical aspects of the use of data should also be considered within the methodological and strategic perspectives above.

Two issues are now discussed as examples of the ethical implications of this review on the use of data to support learning. The first issue regarding the existence of ethical guidelines leads to the other – an optimistic outlook about the education community’s ability to minimise the negative consequences for students of the use of data.

Ethical guidelines

The Australasian Curriculum Assessment Certification Authorities (ACACA) is the national body for the chief executives of the statutory bodies in the Australian States and Territories and in New Zealand responsible for certificates of senior secondary education. In an ACACA publication, *ACACA Guidelines for Assessment Quality and Equity* (1997), equity is taken to be a synonym for fairness and it is stated that fairness and quality are closely related. It can be deduced then that these guidelines encourage the exercise of integrity and rigour, as called for earlier.

Since the ACACA agencies are responsible for senior certification around the country, they hold large, rich, quality-assured datasets about students and student achievement. In another of their publications, *ACACA Data Release and Presentation Guidelines* (2005), a statement is made about the ethical issues associated with the release of datasets about senior student performance (the sort of release that makes it easy for newspapers to compile league tables).

The principles [regarding ethical release] involve recognition that the issues are in essence ethical – we acknowledge our responsibility to our communities for how we select and present data.

(ACACA, 2005, p. 1)

Elsewhere the distinction is made between the datasets held by the statutory authorities and those from well-designed research studies (ACACA, 2005). Cautionary notes are sounded about using the data in misleading ways, which might harm individuals and organisations. The obvious example of such misuse is a statement like, ‘This school is more successful than that school’, which, as discussed in Section 4, might be a completely incorrect piece of information.

Optimism

Notwithstanding the sort of ethical issues acknowledged above, which are also addressed by agencies and individual operators worldwide, consequences of the use of data for individual students or groups of students can still turn out to be negative, regardless of whether those data come from high-stakes assessments or low-stakes assessments, from questionnaires about teacher practice or surveys about the health of a system. While this review has provided examples of dubious public treatment of student achievement data, conversations during the ACER 2005 Research Conference and the content of the papers delivered at that conference engendered optimism from the ethical perspective for the following reasons: participants and papers reinforced the primacy of student learning in the assessment process; they supported the promulgation, monitoring and maintenance of standards; and they acknowledged that

educational assessment is a human activity and therefore not values-free. Such conversations and presentation content are markers of integrity and rigour. Thus, we are in a position to minimise the negative consequences for systems, schools, teachers and, most importantly, students.

Summary comments on ethical perspective

In the summary of the methodological perspective, there were reminders of the current environment of increased use of student performance data for accountability purposes. In the summary of the strategic perspective, there were reminders of the consequences for teachers and students of evidence–practice gaps. Earlier in this discussion of the ethical perspective, there was a reminder (ACACA, 1997) that equity and quality are inextricably linked. In the never-ending quest for fairness for all in the education delivery system, it is therefore timely to note two current phenomena. One is the constant vigilance of the education community to the negative effects of data use in the learning environment. The other is the constant efforts of the kind currently in evidence in Australia and overseas to promote high quality in all aspects of education by using data to support learning.

Knowledge is not a loose-leaf notebook of facts. Above all, it is a responsibility for the integrity of what we are, primarily of what we are as ethical creatures.

(Bronowski, 1973, p. 436)

Conclusion

In looking at implications for practice, policy making, and program implementation and reform in Australia of what has gone before in this review paper, three perspectives were examined in Section 5 – the methodological, the strategic, and the ethical. The theoretical components of these perspectives have been made concrete by a few practical examples. Short discussions about professional development, teachers’ attitudes and expertise, policy makers and data, evidence–practice gaps, ethical guidelines, and an optimistic view of fair play in the use of data to support learning have given further substance to the words and meaning of the conference theme ‘using data to support learning’.



Concluding comments

This review has addressed issues identified and discussed at the ACER 2005 Research Conference and through an eclectic assemblage of published literature in the educational field (particularly assessment, accountability, and reform), as well as in fields outside education which complement the use of data to support learning.

The question of what constitutes useful evidence recurred in the review's examination of the use of data in supporting learning. It has been re-asserted throughout that all data use needs to be explicitly guided by documented analysis and rigorous discussion of the reliability of the data and the validity of the interpretations.

Based on the identified issues, and the analytical text in Section 4, the review advocates that eight recommendations are considered by those working in the field. They present themselves in four clusters.

The first cluster of suggestions can be envisaged as one-off projects to be undertaken by experienced researchers. Each of them goes to the core of what is valued in assessment – or at what is seen to be valued, and reflect areas of international interest which are currently very relevant in this country. One of the suggested projects is the collation of, and reflection on, the existing body of cutting-edge research on using student work (including products of extended computer-based assessments) as a data source. The second suggested project is a detailed review of the attitudes of Australian teachers and policy makers to the application of 'assessment-for-learning'.

The second cluster of two suggestions can be envisioned as necessary and important professional development activities for the professional learning community, activities based on the best evidence about professional development models. The rationale for them is embedded in the intricacies of previous discussions about using data (by whom? about whom?) to support learning (whose learning?). The first of the two suggested activities is hands-on training for system administrators in the use of large-scale datasets, with special emphasis on using these datasets in formulating educational policy. The second suggestion is for the provision of professional development programs for teachers in techniques for interrogating student data, especially data supplied to schools by external agencies and assessment data generated at the school level.

The third cluster relates to the most serious advocacy, which is for two research studies that should be commissioned. The first is that a rigorous appraisal of existing research findings related to current national and international issues by a coalition of educational interests be undertaken and published. This may appear to be a rather general kind of suggestion but the development of a research-based context for future reports would not only be most informative

of itself, but it would also assist those who, when confronted by reports, are inclined to rely on the executive summaries and media releases circulated by researchers and policy makers. The media 'take' on research and public discussion of it may be improved. The second proposed research study, which is specific to the topic of this review and directly related to the conclusion in Section 3, is for the conduct of a meta-analysis of the use of data to support learning. The questions to be asked are: Is it generally the case that data are used to support learning? In what ways? If or when data are used, is their use accompanied by a demonstrable improvement in learning?

The fourth cluster contains two suggestions that revolve around the issue of ways of operating. There is evidence to suggest that, in the current general climate and professional learning environment, it is critical that there be a commitment to a multi-disciplinary approach when briefs are prepared for practitioners and policy makers on educational research findings about a pertinent issue. The second suggestion may seem obvious but it is nevertheless one that cannot be excluded from moving forward in using data to support learning: it is that there should be better use of research findings at all levels of the education delivery system in this country. If the strategies suggested in the second cluster of suggestions from this review had been implemented, it would be much easier for this outcome to be achieved. Currently there is a huge volume of valuable information that is not being tapped, shared, reflected upon, debated, and exploited as fully as it could be, between researchers and bureaucracy, and between systems and teacher practitioners, in the service of supporting learning.

Finally, while none of the above suggestions or recommendations is especially earth-shattering, the effects on the profession of their being done properly, based on evidence of what works, could be profound. To assist in achieving them, this review has provided ideas to provoke further debate, practical hints to be incorporated into professional practice, descriptions of experiences and research findings to inform policy making and program implementation, and prompts for further discussions amongst classroom practitioners, whole-school stakeholders, parents, and system bureaucrats.

List of 2005 ACER Research Conference papers

At the conference, 3 keynote papers (Kingsbury, Earl and Hattie), 18 concurrent papers and ACER CEO Masters' opening and closing addresses were presented.

Abstracts of these 23 papers are available for downloading on the conference website.

The link to that website is: <http://www.acer.edu.au/workshops/conferences.html>

Allen, J. R. *Using the evidence of student achievement for improvements at individual, class and school level.*

Angelico, T. *An evidence-based approach to improvement: A case study of the Victorian Catholic sector.*

Axworthy, D. *Turning data into information that improves learning: The WA experience.*

Bruniges, M. *An evidence-based approach to teaching and learning.*

Cahill, R. *Getting It Right ... using the right data effectively.*

Craig, W. *Data and school improvement – A school perspective.*

DeCourcy, J. *Using HSC data to give principals leverage.*

Earl, L. M.¹ *From accounting to accountability: Harnessing data for school improvement.*

Hattie, J. A. C. *What is the nature of evidence that makes a difference to learning?*

Hollingsworth, H. *Learning about teaching and teaching about learning: Using video data for research and professional development.*

Holmes-Smith, P. *Assessment for learning: Using statewide literacy and numeracy tests as diagnostic tools.*

Ingvarson, L. *Getting professional development right.*

Kingsbury, G. G. *Benchmarks and growth and success ... Oh, my!*

Matters, G. N. *Good data, bad news, good policy making ...*

Masters, G. N. Opening address (ppt slides): *The role of information in professional work.*

Masters, G. N. Closing address (ppt slides): *A few threads ...*

Meiers, M. *Evaluation of the Getting it Right Literacy and Numeracy Strategy in Western Australian schools.*

Richardson, C. *Data-informed research and practice: Evaluating student achievement in secondary schools.*

Rowe, K. *Evidence for the kinds of feedback data that support both student and teacher learning.*

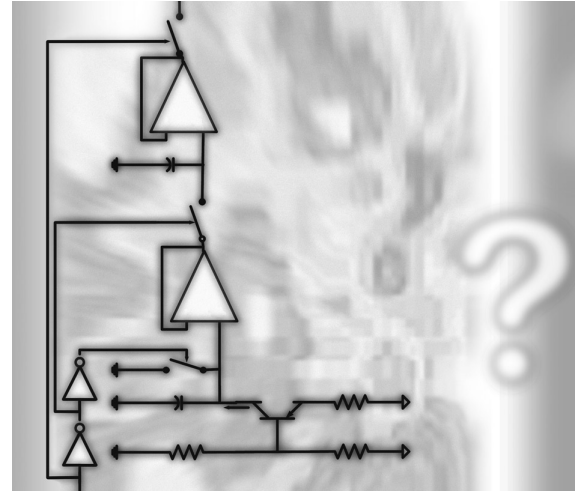
Rowley, G., & Congdon, P. *Data-driven school improvement through the VCE Data Service.*

Smith, M. *Getting SMART with data in schools: Lessons from NSW*

Tognolini, J. *Using online assessment to inform teaching and learning in primary and secondary schools.*

Tozer, L., & Holmes, M. *Moving on from Count Me In Too: Evidence-based teaching and learning in numeracy in the early and middle year of schooling.*

¹ This paper is referred to as Earl (2005a) in the review.



References

- Antman, E. M., Lau, J., Kupelnick, B., Mosteller, F., & Chalmers, T. C. (1992). A comparison of results of meta-analyses of randomised control trials and recommendations of clinical experts. *Journal of the American Medical Association*, 268, 240–248.
- Assessment Reform Group (ARG). (1999). *Assessment for learning: Beyond the black box*. Cambridge: University of Cambridge, School of Education.
- Australasian Curriculum Assessment Certification Authorities (ACACA). (1997). *Guidelines for assessment quality and equity*. Brisbane: Queensland Board of Senior Secondary School Studies.
- Australasian Curriculum Assessment Certification Authorities (ACACA). (2005). *Guidelines for data release and presentation*. 23/12/2005, retrieved from <http://www.acaca.org.au/documents.htm>
- Baker, E. L., O'Neil, H. F., & Linn, R. L. (1993). Policy and validity prospects for performance-based assessment. *American Psychologist*, 48(12), 1210–1218.
- Bennett, D. C. (2001). Assessing quality in higher education. *Liberal Education*, 87, 40.
- Bennett, R. E. (2000). *How the Internet will help large-scale assessment reinvent itself*. Paper presented at the 36th IAEA conference, Jerusalem.
- Biggs, J. (1999). *Teaching for quality learning at university*. Buckingham: SRHE & Open University Press.
- Biggs, J., & Moore, P. J. (1993). *The process of learning*. 3rd ed. New York: Prentice Hall.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.
- Boomer, G. (1988). *Garth Boomer in conversation with Charles Galloway and Judith Green*. Unpublished manuscript, October 26, 1988.
- Bowe, R., & Ball, S. J., with Gold, A. (1992). *Reforming education and changing schools: Case studies in policy sociology*. London: Routledge.
- Bronowski, J. (1973). *The ascent of man*. London: British Broadcasting Corporation.
- Brown, G. T. L. (2002). *Teachers' conceptions of assessment*. Unpublished doctoral dissertation, Auckland: University of Auckland.
- Burke, K. (18 Dec. 2004). Class act: HSC results best yet. *The Sydney Morning Herald*.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, 25(2), 119.
- Craig, J. C., Irwin, L. M. and Stockler, M. R. (2001). *Medical Journal of Australia*, 174, 248–253.

- Dahl, R. A. (1970). *Who governs? Democracy and power in an American city*. New Haven, CT: Yale University Press.
- David, H. A. (1987). Ranking from unbalanced pair-comparison data. *Biometrika*, 74, 432–436.
- Department of Education and the Arts. (2004). *The New Basics Research Report*. Brisbane: Assessment & New Basics Branch.
- Doherty, L., Norris, J., & Burke, K. (17 Dec. 2005). State school blitz of top HSC spots. *The Sydney Morning Herald*.
- Earl, L. M. (2005b). *Assessment as learning: Using classroom assessment to maximise student learning*. Thousand Oaks, CA: Corwin Press Inc.
- Eisner, E. W. (2001). What does it mean to say a school is doing well? *Phi Delta Kappan*, 82(5), 348–372.
- Ellis, L. A. (2005). *Balancing approaches: Revisiting the educational psychology research on teaching students with learning difficulties* (Australian Education Review). Camberwell, Victoria: ACER.
- Ericson, D. P., & Ellett, F. S. (2002). The question of the student in educational reform. *Education Policy Analysis Archives*, 10(31). 24/10/2005, retrieved from <http://epaa.asu.edu/epaa/v10n31/>
- Forster, M. (2001). Accepting the challenge: Assessing the moral and ethical outcomes of schooling. In S. Pascoe (Ed.) (2002). *Values in Education. Australian College of Educators, College Year Book 2002*. Canberra: Australian College of Educators.
- Forster, M. (2005). *Reporting for learning, Reporting as learning*. Paper presented at 10th Annual National Roundtable Conference, Melbourne, 24–25 October 2005.
- Friend, L. M. (2002). *Examining the construction of knowledge at local sites in New Basics schools*. Report to the Assessment & New Basics Branch, Queensland Department of Education and the Arts, Brisbane.
- Freebody, P. R. (2005a). *Turning 'alien beings into robust trustees of our resources and ideals': New times for the senior phase*. Paper presented to the Review of Syllabuses for the Senior Phase of Learning, Queensland Studies Authority, Brisbane, 3–4 October 2005.
- Freebody, P. R. (2005b). *Background, rationale and specifications: Queensland Curriculum Assessment and Reporting Framework*. Brisbane: Department of Education and the Arts.
- Fullan, M. G. (1993). *Change forces: Probing the depths of educational reform*. New York: Falmer Press.
- Fullan, M. G. (1999). *Change forces: The sequel*. New York: Falmer Press.
- Fullan, M. G. (2000). The three stories of education reform. *Phi Delta Kappan*, 81(8), 581–584.
- Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York: Basic Books.
- Gipps, C. for Education Queensland's Assessment & Reporting Taskforce. (2002). *The Report of the Assessment and Reporting Taskforce*. Brisbane: Department of Education.
- Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Buckingham: Open University Press.
- Goldstein, H. (1997). Methods in school effectiveness research. *School Effectiveness and School Improvement*, 8(4), 369–395.
- Goldstein, H. (2005, in train). International comparisons of student attainment: Some issues arising from the PISA study. *Assessment in Education* (to appear).

- Goldstein, H., & Spiegelhalter, D. J. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society Series A*, 159, 385–443.
- Goldstein, H., & Woodhouse G. (2000). School effectiveness research and educational policy (1). *Oxford Review of Education*, 26(3), 353–363.
- Hargreaves, D. (1996). *Teaching as a research-based profession: Possibilities and prospects*. Teacher Training Agency Annual Lecture 1996. London: Teacher Training Agency.
- Hawley, W., & Valli, L. (1999). The essentials of effective professional development: A new consensus. In L. Darling-Hammond, & G. Sykes (Eds.) (1999). *Teaching as the learning profession. Handbook of policy and practice* (127–150). San Francisco: Jossey-Bass Publishers.
- Kaplan, A. (1964). *The conduct of inquiry*. San Francisco: Chandler.
- Killion, J., & Bellamy, G. T. (2000). On the job: Data analysts focus school improvement efforts. *Journal of Staff Development*, 21(1), 12–18.
- Kimmelman, P. L., & Kroeze, D. J. (2002). *Achieving world class schools: Mastering school improvement using a genetic model*. Norwood, Mass: Christopher-Gordon Publishers.
- Kingsbury, G. G., & Houser, R. (1997). Using data from a level testing system to change a school district. In *The Rasch tiger ten years later: Using IRT techniques to measure achievement in schools*. Chicago: National Association of Test Directors.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Linn, R. L. (Ed.) (1989). *Educational Measurement 3rd edition*. Washington, DC: The American Council on Education and The National Council on Measurement in Education.
- Luke, A. (2000). *New Basics Project Technical Paper*. Brisbane: Department of Education.
- Masters, G. N., Meiers, M., & Rowe, K. J. (2003). Understanding and monitoring children's growth. *Educare News*, 136, 52–53.
- McGaw, B. (2002). *Raising the bar and reducing the failures: A possible dream*. Plenary address. ACER 2002 Research Conference, Sydney.
- McGaw, B. (2005). *Achievement and equity: Where is Australia at?* Keynote address, Curriculum Corporation 12th National Conference, Brisbane.
- Mitroff, I. I., & Sagasti, F. (1973). Epistemology as general system theory: An approach to the design of complex decision-making experiments. *Philosophy of Social Sciences*, 3, 117–134.
- Moss, P. A. (1994). *Can there be validity without reliability?* *Educational Researcher*, 23(2), 5–12.
- Nyborg, H. (2003). The sociology of psychometric and biobehavioral sciences: A case study of destructive social reductionism and collective fraud in 20th century academia. In H. Nyborg (Ed.). *The scientific study of general intelligence* (441–501). Amsterdam: Elsevier Science/Pergamon.
- OECD. (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: Organisation for Economic Co-operation and Development.
- Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan*, 68, 679–682.
- Richardson, C. (2003). *Fairer methods of assessment: Value-added analyses of final secondary results*. Paper presented at the 15th International Congress of School Effectiveness and Improvement, Sydney.
- Rowntree, D. (1981). *Statistics without tears: A primer for non-mathematicians*. Ringwood, Vic: Penguin Books Australia.
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy and Practice*, 5, 77–84.

- Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee value-added assessment system, a quantitative, outcomes-based approach to educational measurement. In Jason Millman (Ed.). *Grading teachers, grading schools, Is student achievement a valid evaluation measure?* (137--162). Thousand Oaks: CA Corwin Press.
- Scott, I., Heyworth, R., & Fairweather, P. (2000). The use of evidence-based medicine in the practice of consultant physicians. Results of a questionnaire survey. *Aust N Z J Med* 2000, 30, 319-326.
- Schlipp, P. A. (Ed.) (1949). *Albert Einstein: Philosopher-Scientist*. New York, Tudor.
- Seashore Louis, K., Febey, K., & Schroeder, R. (2005). State-mandated accountability in high schools: Teachers' interpretations of a new era. *Educational Evaluation and Policy Analysis*, 27 (2), 177.
- Seegerstrale, U. (2000). *Defenders of the truth: The battle for science in the socio-biology debate and beyond*. New York: Oxford University Press.
- Shepard, L. (1991). Psychometrician's beliefs about learning. *Educational Researcher*, 20(7), 2-16.
- Shulman, L. S. (1988). Disciplines of inquiry in education: An overview. In R. M. Jaeger (Ed.). *Complementary methods for research in education* (3-17). Washington, DC: American Educational Research Association.
- Stage, C. (1994). *Gender differences on the SweSAT: A review of studies since 1975*. Department of Educational Measurement, Umeå University, EM No. 7.
- Stenhouse, L. (1975). *An introduction to curriculum research and development*. London: Heinemann Educational.
- Stenhouse, L. (1981). What counts as research? *British Journal of Educational Studies*, 29(2).
- Stobart, G. (2004). *The formative use of summative assessment: Possibilities and limits*. Paper presented at the 30th IAEA Conference, Philadelphia.
- The University of Queensland, School of Education. (2001). *The Queensland School Reform Longitudinal Study*. Brisbane: Department of Education.
- Thomas, B. (22 Jan. 2006). In 'Good mail'. *The Sunday Mail*, p. 57.
- Thorndike, R. L. (Ed.). (1971). *Educational Measurement 2nd edition*. Washington, DC: American Council on Education.
- US Government. (2001). *No Child Left Behind (NCLB) Act*.
- Victorian Department of Education & Training (2005). *Interpreting the 2005 attitudes to school survey reports v1.1*. Melbourne, Victoria: Author.
- Wainwright, M. A., Wright, M. J., Luciano, M., Geffen, G. M., Martin, N. G. (2005). Multivariate genetic analysis of academic skills of the Queensland Core Skills Test and IQ highlight the importance of genetic *g*. *Twin Research and Human Genetics*, 8(6), 602-608.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92, 548-73.
- Wiliam, P. (2004). *Keeping learning on track: Integrating assessment with instruction*. Paper presented at closing session, 30th IAEA Conference, Philadelphia.
- Willingham, W. W., & Cole, N. S. (Eds.). (1997). *Gender and fair assessment*. Princeton, NJ: Lawrence Erlbaum Associates.
- Wilson, E. O. (1998). *Consilience: The unity of knowledge*. New York: Random House.